An Efficacy Study of READ 180

A PRINT AND ELECTRONIC ADAPTIVE INTERVENTION PROGRAM GRADES 4 AND ABOVE

Council of Great City Schools

Final Report

Study of *READ 180* in
the Council of Great City Schools

Prepared by

Interactive, Inc.
January 30, 2002

READ 180 Validation Study: FINAL REPORT
Table of Contents

List of Tables

List of Figures

I. EXECUTIVE SUMMARY

Scholastic, Inc. is committed to helping all students become lifelong readers. Scholastic has therefore created the *READ 180* program, which is intended to help low performing students along the path toward literacy success. *READ 180* combines research-based practices of reading instruction with the most effective use of technology in the classroom and is based upon a prototype called the Peabody Learning Lab, developed at Vanderbilt University.

Interactive, Inc. has been contracted to conduct an independent validation study of the effects of *READ 180* on the reading achievement of low performing students in seven member districts of the Council of Great City Schools. These districts are:

Atlanta, GA; Boston, MA; Columbus, OH; Dallas, TX; Houston, TX; Miami-Dade, FL; and San Francisco, CA. This report presents findings, conclusions, and recommendations from the study.

In evaluating the overall effectiveness of the *READ 180* program in this study, we address the following research questions:

1. What impact does *READ 180* have on student reading achievement and reading proficiency?
2. How faithfully did teachers implement *READ 180* and what factors mediate the level of implementation?
3. How is fidelity of implementation related to various student outcomes? i.e., Does more faithful implementation of *READ 180* result in greater student outcomes?

The summary responses to those questions are as follows:

1. Fidelity of Implementation:
   As predicted by the research on school program implementation, schools forced more adaptation from the program to the school than *vice versa.* That is to say, more often than not, schools changed the processes (e.g., scheduling, technological infrastructure, etc.). Even within those structurally-imposed modifications, however, implementation within individual classrooms was generally of high quality. All but one of the teachers involved in the validation study indicated that they would recommend *READ 180* to other middle school teachers.

2. Achievement Gains:
   Across all *READ 180* students in the three districts that provided us with year-to-year test scores (Boston, Dallas and Houston), the difference in the growth on the Stanford-9 (Total Reading) between the treatment (+22.94) and control groups (+17.24) is statistically significant (F1 2.624, a = .000) and in favor of the students in the *READ 180* classes. Further, using analysis of covariance to control for prior levels of achievement, the difference in the adjusted mean between the treatment (648.48) and control groups (642.42) is statistically significant (F=1 5.396, a = .000) and in favor of the students in the *READ 180* classes.

   Since each standard deviation gain is approximately equal to gains of an academic year of 10 months, each tenth of a standard deviation can be viewed as roughly 1 month of achievement gain. Since the standard deviation of the scaled scores of achievement tests are different for each level of the test, separate effect size calculations must be done for each grade level cohort. Disaggregating the differences by grade level, and standardizing the differences, we find effect sizes of .26 standard deviations at 6th grade and .43 standard deviations at 8th grade. Translated to the academic year, **the 6th grade *READ 180* students experienced roughly 2.5 months more growth in reading achievement than other similar students, and the 8th grade *READ 180* students experienced over four months more growth than their control group counterparts.**

   In Columbus, where pretests were administered in the fall of academic year 2000-01, the difference in the growth on the reading comprehension subtest of the Stanford-9 between the treatment (+14.41) and control groups (-3.77) is statistically significant (F=13.182= .000) and in favor of the students in the *READ 180* classes. The differences in growth in Columbus rise to the level of considerable educational significance.

3. Extent of Implementation and Extent of Achievement Gain:
   Control group students, i.e. students with no exposure to *READ 180,* did not exhibit the same kind of positive academic progress demonstrated by students in classrooms where implementation quality was classified as either moderate or high.

This validation study tested *READ 180* under world-of-practice conditions. The research design was necessarily modified and tailored to document the experiences with *READ 180* in the districts and schools as they chose to implement it. Researchers did not control the teaching and learning process. That is as we believe it should be. Despite variation in practice, *READ 180* still added value to the districts, schools, teachers and students and that testifies to the strength of the program.

II. INTRODUCTION: *READ 180* AT THE INTERSECTION OF LITERACY INSTRUCTION AND EDUCATION REFORM

A. MIDDLE-SCHOOL LITERACY INSTRUCTION

> Adolescents becoming adults in the 21st century will read and write more than at any other time in human history. They will need advanced levels of literacy to perform their jobs, run their households, act as citizens, and conduct their personal lives. They will need literacy to cope with the flood of information. They will need literacy to feed their imaginations so they can create the world of the future. In a complex and sometimes even dangerous world, their ability to read will be crucial. Continual instruction beyond the early grades is needed.[1]

Although middle school educators recognize literacy growth and development as important, many schools attended by adolescents do *not* include reading instruction in the curriculum for all students. Middle school programs often emphasize a broad set of student needs including physical, emotional, and social, as well as academic. Language arts teachers are held accountable for covering a literature program without consideration of literacy levels, growth or development.

The results of this neglect of literacy development in the middle grades are now well documented. On the 1998 National Assessment of Educational Progress (NAEP),[2] one out of every four eighth-grade students did not read well enough to be considered above the Basic level. Further, only 33% of eighth-grade students performed above the Proficient level. Two out of every three eighth-graders are not "proficient" readers.

Yet, despite the NAEP data, very little literacy development or intervention is provided in middle schools.

At their January 1997 meeting, the Board of Directors of the International Reading Association (IRA) approved the formation of the Commission on Adolescent Literacy ("Commission"). The Commission was given a 3-year opportunity to advise the Board on policies and priorities related to the literacy of adolescents.

---

[1] Moore, D.W., Bean, T.W., Birdy shaw, D., & Ryckik, J.A. (1999) Adolescent *Literacy: A Position Statement.* Newark, DE: International Reading Association.

[2] The nation's only ongoing survey of what students know and can do in various academic subject areas. NAEP is authorized by Congress and administered by the National Center for Education Statistics in the Department of Education. In 1998, NAEP conducted a national reading assessment of fourth-, eighth-, and twelfth-grade students, and a state-by-state reading assessment of fourth- and eighth-grade students.

In March of 1999, the Commission proffered a position statement on adolescent literacy. The theme of the statement is that "adolescents deserve more" with respect to the teaching and development of literacy. In that statement, the Commission established the seven "Principles for Supporting Adolescents' Literacy Growth" outlined in the following table.

Table 1 Principles for Supporting Adolescents' Literacy Growth

1. Adolescents deserve access to a wide variety of reading material that they can and want to read.
2. Adolescents deserve instruction that builds both the skill and desire to read increasingly complex materials.
3. Adolescents deserve assessment that shows them their strengths as well as their needs and that guides their teachers to design instruction that will help them grow as readers.
4. Adolescents deserve expert teachers who model and provide explicit instructions in reading comprehension and study strategies across the curriculum.
5. Adolescents deserve reading specialists who assist individual students having difficulty learning how to read.
6. Adolescents deserve teachers who understand the complexities of individual adolescent readers, respect their differences, and respond to their characteristics.
7 Adolescents deserve homes, communities, and a nation that will support their efforts to achieve advanced levels of literacy and provide the support necessary for them to succeed.

B. *READ 180* as a Building Block for Comprehensive School Reform

Scholastic, Inc., in its commitment especially to the first three of these principles, created *READ 180* "to help struggling students break out of their failure cycles and turn their lives around 180 degrees." *READ 180* is based on a prototype, Peabody Learning Lab, developed at Vanderbilt University. *READ 180* combines research-based practices of reading instruction with the computer-related effective use of technology in the classroom.

READ 180, however, is more than just a literacy program. Rather, it is a comprehensive, multi-faceted instructional model along the lines of Comprehensive School Reform Demonstration (CSRD) programs. In an effort to aid the states in education reform, the United States Congress established Comprehensive School Reform Demonstration (CS RD) in the Fiscal Year 1998 appropriations legislation for the U.S. Department of Education (Public Law 105-78). Basically, the CSRD Program provides financial incentives for schools to implement comprehensive school reform programs that are based on research and effective practices and that include an emphasis on academics and parental involvement. Federal funds are distributed to the states that, in turn, fund schools that are willing and able to implement whole-school reform. Schools that receive funds are expected to plan and implement programs that deploy, in a coherent fashion, the following nine components (as specified in the law):

- Effective, research-based, replicable methods and strategies
- Comprehensive design with aligned components
- Professional development
- Measurable goals and benchmarks
- Support within a school
- Parental and community involvement
- External technical support and assistance
- Evaluation strategies, and
- Coordination of resources.

The CSRD legislation names 17 models as examples only. In addition, while schools or districts can develop their own models, the Northwest Regional Educational Laboratory maintains a catalog of models for comprehensive school reform. The most current catalog lists models divided into three categories: (1) Entire-school Models, (2) Subject Area Models (reading/language arts, mathematics and science), and (3) Other Models (those that do not fit neatly into either of the other two categories). Entire-school models provide schools with a framework for change covering most or all aspects of school operations, while subject area and other models can be used as building blocks for comprehensive reform. The “other” models include thinking skills models, classroom management models, and behavioral skills models.

C. THE *READ 180* INSTRUCTIONAL MODEL

READ 180 qualifies as a subject-area model and therefore a building block for school reform. The *READ 180* Instructional Model consists of a 90-minute literacy block. During that block, small classes of 15-18 students spend the first 10 minutes together with the teacher doing language arts instruction. Over the next hour, the class breaks into three smaller groups, which proceed through three 20-minute rotations as follows:

- ✔ **Small-group Instruction**
  The *READ 180* teacher sits with 5-6 students doing group reading and/or language arts instruction.

- ✔ **Independent Reading**
  The students retreat to **a comfortable area where** they read leveled[3] paperbacks with the option of adding audio through headphones as modeled reading.

- ✔ **Direct Instruction (computers)**
  Nine topical **CD-ROMs** provide students with background knowledge and mental models through full-motion video. Students then encounter a reading passage based on the video that is at the appropriate ability-level as determined by an electronic placement test at the beginning of the program. After the video and the reading passage, students proceed through the following three zones on each Topic CD:

  - ➢ **Word Zone**
    Instruction for developing basic decoding skills and the ability to use structural analysis
  - ➢ **Spelling Zone**
    Extensive individualized instruction in the acquisition and transfer of spelling patterns and sounds
  - ➢ **Success Zone**
    Students are assessed for comprehension, word recognition and fluency skills.

In sum, *READ 180* is a multi-faceted and comprehensive middle-school literacy program that targets the neediest students. The program is based on extensive research on how young people learn to read, and it combines varying pedagogical styles (whole-class, direct instruction and small-group rotations), rich, relevant texts, and contemporary learning technology. It is comprehensive, yes, but it is also different. It is nearly impossible to visit a well-implemented *READ 180* classroom and not recognize that teaching and learning is done differently there. In other words, *READ 180* is both comprehensive and innovative. The remainder of this report demonstrates that *READ 180* is also very valuable.

---

[3] That is, content geared to their developmental stage.

III. METHODS

The Validation Study

During the 1998-99 academic year, Scholastic, Inc. entered into a partnership with the Council of Great City Schools. Seven of the largest, urban school districts would have the opportunity to implement *READ 180* in middle schools attended by some of the lowest functioning adolescent readers. Seven of the Great City School Districts agreed to implement *READ 180* and be part of this validation study. Those districts include:

- Atlanta, GA
- Boston, MA
- Columbus, OH[4]
- Dallas, TX
- Houston, TX
- Miami-Dade, FL
- San Francisco, CA

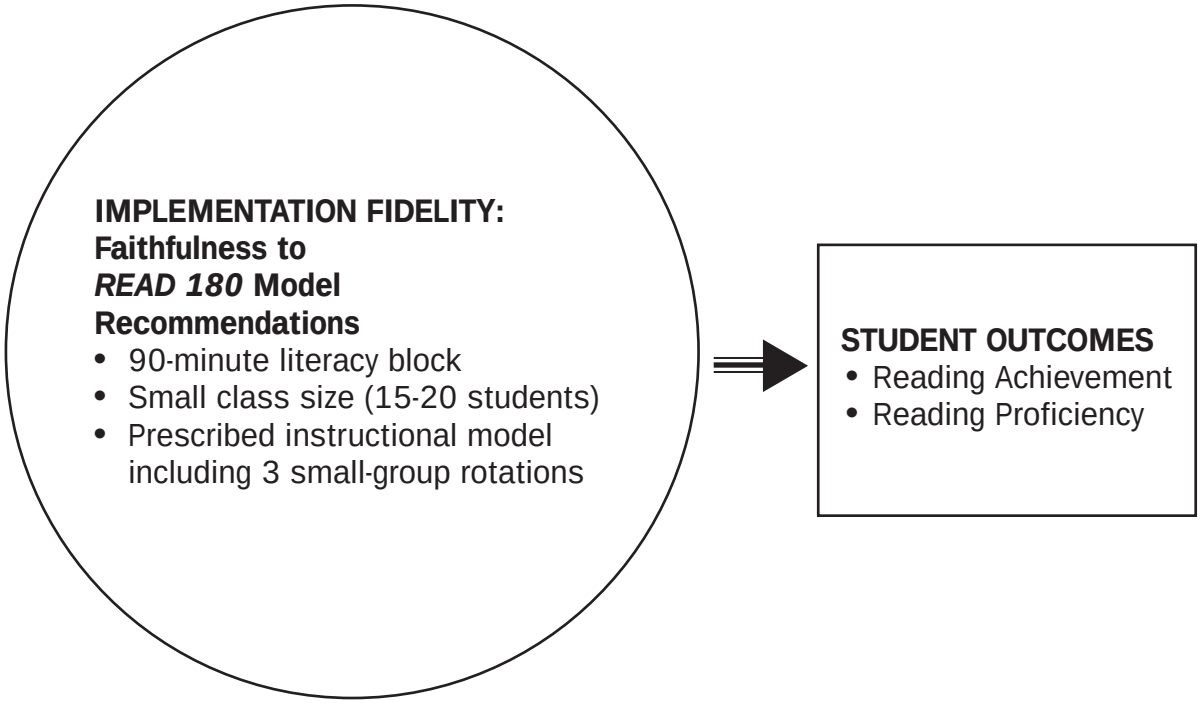
A. Analytic Framework

As with any school program, outcomes are necessarily (but not only) a function of implementation fidelity. That is to say, educational "programs" can effect changes in educational outcomes (achievement, attendance, behavior, etc.) only to the extent that the program is implemented faithfully.

[4] Chicago was originally included as part of the validation study. However, instructional policies in Chicago conflicted with the demands of the original research design. Therefore, Columbus was chosen to replace Chicago.

Therefore, the research design for the *READ 180* Validation Study is based on the model depicted in Figure 1.

Figure 1. Research Design Model



READ 180 is not a new textbook; neither is it just a piece of software. *READ 180* is a comprehensive, sophisticated, research-based literacy program designed around the fundamental instructional model described above. We describe issues of implementation fidelity more fully in section IV of this report. The student outcomes of interest are those domains targeted for improvement by *READ 180*: student reading achievement and reading proficiency.

Thus, in evaluating the overall effectiveness of the *READ 180* program in this study, we address the following research questions:

1. How faithfully did teachers implement *READ 180* and what factors mediate the level of implementation?
2. What impact does *READ 180* have on student reading achievement and reading proficiency?
3. How is fidelity of implementation related to various student outcomes? i.e., Does more faithful implementation of *READ 180* result in greater student outcomes?

B. Research Design

1. The Basic Research Design

At the most general level, the validation study followed a randomized control group, pretest-posttest design.

Table 2. Research Design

	Pretest	Treatment	Posttest
Treatment Group	T1	X1	T2
Control Group	T1		T2

The research design agreed upon by Scholastic, The Council of Great City Schools, and the validation districts called for each of the seven districts to be responsible for selecting and managing their own samples. Each district agreed to recruit two middle schools to the *READ 180* program and validation study. Each middle school was to establish two *READ 180* classrooms to serve a total of 120 students (four classes of 15 students in each of the two classrooms). Therefore, the expected sample size was 1,680 *READ 180* students (7 districts x 2 schools x 120 *READ 180* students[5]).

In addition, each validation school was to rank order their 6th and/or 7th" grade students by reading ability. The lowest 320 students were to be randomly assigned to a *READ* 180 class (n=120), the control group (n=120), or the backup group (n=80),[6] but no students with a reading Grade Equivalent lower than 1.5 were to be placed in *READ 180* classes.[7]

Finally, the design called for the validation districts to administer a common norm-referenced, standardized achievement test (the Stanford-9) to all the students in the *READ 180* program (at T1, T2 and T3 in the graphic above). At the very least, district officials agreed to administer the multiple-choice reading subtests and the open-ended reading assessment. The Stanford-9 was to serve as a common metric for student reading achievement.

---

[5] Research directors for each of the seven validation districts agreed to this sampling framework in the Spring of 1999. By the time AY 2000-01 commenced, districts implemented *READ 180* in a variety of ways with varying numbers of students.

[6] High mobility rates are typical in the population of the validation districts, so backup students were to serve as replacements for *READ 180* students who left the program for any reason. Further, there was a distinct possibility that a number of students originally placed into the *READ 180* program would reach grade level at some point over the course of the academic year. Those students were to be transitioned into a regular classroom and replaced by students in the backup group. Therefore, the plan was to administer all pretest instruments to the backup students so baseline data would be available for them if they were to enter the program.

[7] *READ 180* is designed as a middle school reading intervention program. Therefore, the material would be too advanced for 6th or 7th' grade students reading at a grade equivalent of 1.5 or lower.

2. A Note About the Control Groups

This validation study was designed to include control groups comprised of students designated as such through random assignment at the level of the individual student. However, in most cases, control group assignment did not proceed accordingly. Often, and rightfully so, administrative and/or pedagogical needs trumped the demands of clinical or experimental research.

For instance, a number of the schools decided that there was a group of students for whom assignment to the *READ 180* classroom would be most beneficial. Only after those students were assigned to *READ 180* was a reasonably comparative control group identified. In addition to those sorts of deviations from the random assignment plan, there were instances where individual parents or caregivers requested inclusion or exclusion from the program. Also, there are cases of students who asked out of the program. In most of these cases where individual students opted out of *READ 180*, the student was replaced, but the fact remains that there was rarely, if ever, true random assignment.

In addition to the loss of true random assignment, the realities of local control confounded our ability to completely understand the curricular and instructional practices to which the control groups were exposed. That is to say, our purpose for including a treatment group and a control group in the study was to be able to compare *READ 180* to the status quo in the district or school. However, it is not the case that the students assigned to the control group all experienced the same literacy intervention curriculum. Nor is it the case that control group students within the same district, or even the same school, experienced a common curriculum.

As one example, the Houston Independent School District recently completed an audit of their middle-school reading curricula and found between 50 and 60 different programs in place in schools across the district. There is no standard, district-wide middle-school reading program. Therefore, the particular reading program that students in the control group experienced was dependent on the classroom or teacher to which the students were assigned.

In the Columbus, Ohio public schools, the *READ 180* students in the validation schools were the lowest-performing students in the schools, so control groups from those same schools would be non-equivalent (i.e. slightly higher on the pretest). Therefore, students from three non-*READ 180* schools were assigned to the control groups. Furthermore, the district operates a “Safety Net” program for students who perform at low levels on reading proficiency tests. Under the general umbrella of the Safety Net program, each school in the district with a significant number of under-performing students can choose from a series of literacy interventions to be implemented in the school. Students in the Columbus control group attended three different schools and, therefore, experienced as many as three different literacy interventions other than *READ 180*.

This local control phenomenon is fairly consistent across all the validation districts. While it complicates the research design and analysis, it speaks to the inherent difficulties of conducting true clinical trials within the real world of schools. Further, we are confident that we have reasonably equivalent control groups, and where there is non-equivalence we conduct statistical analysis to control for the differences.

3. The Research Design as Fielded

None of the seven districts ultimately followed through with the original research design, despite earlier agreements from district officials in each of the districts. The following table shows some of the research design problems posed by districts.

Table 3. Research Design Concerns

RESEARCH DESIGN FIDELITY ISSUES	
ATLANTA	• In one validation school, a decision was made to keep students in the *READ 180* program for only one-half of the year. • Success for All was implemented district-wide, forcing *READ 180* into a 65-minute block (not 90 minutes). • The district administrator in charge of providing test scores reports significant problems with administration and scoring of the Stanford-9 tests. As a result, we do not yet have test scores from Atlanta.
BOSTON	• In one validation school, teachers set aside 45 of the 90 minutes twice a week to focus on a writing skills curriculum.
COLUMBUS	• Columbus did not administer the Stanford-9 in the Spring prior to the study. As a result, the district administered the multiple-choice assessment in Fall 2000 to serve as the pretest. Thus, Columbus test dates vary from the other districts.
DALLAS	No research-related issues.
HOUSTON	No research-related issues.
MIAMI	• Miami ultimately offered access to only one school for the validation study. • Also, the district repeatedly provided test scores in forms different than requested and agreed.
SAN FRANCISCO	• *READ 180* was ultimately implemented only in one school. Within that one school, only one class of 12 students exists. • We have yet to receive test scores for these students.

Data are analyzed and reported where available.

Very generally, however, to answer the research questions, we collected data as depicted in Table 4.

Table 4. Implementation Measures

IMPLEMENTATION MEASURES	
MEASURE / TECHNIQUE	**CONCEPTS MEASURED**
Teacher Survey: annual	• Teacher attitudes toward *READ 180* • Teacher utilization of various aspects of the program • Teacher perception of student attitudes toward *READ 180*
Site Visits: annual *classroom observations *teacher and principal interviews	• Fidelity of model implementation • Teacher attitudes toward *READ 180* • Teacher utilization of various aspects of the program • Administrator attitudes toward *READ 180*
READ 180 Management Suite Data end of year	• Amount of student exposure to the *READ 180* software.
STUDENT OUTCOME MEASURES	
MEASURE / TECHNIQUE	**CONCEPTS MEASURED**
Standardized Achievement Test: Pre-Post	• Student reading achievement
Scholastic Reading Inventory: Pre-Post	• Student reading proficiency

4. Sample Size

Table 5 and Table 6 show the total number of students for whom we have any piece of data. However, these numbers do not necessarily reflect the reality of how the districts implemented *READ 180.* Further, when the data are analyzed, the n's for the analysis are often significantly lower than the numbers listed in these tables. This attrition occurs for a number of reasons including, but not limited to, the following:

- The district provided us with incomplete test score data on students
- Students were listed in the *READ 180* software management suite as program participants, but no such student exists in the district's test database

Table 5. Total Sample Numbers

Students	Sample Goal	Actual Sample
Treatment	1,680	1,182
Control	1,680	888
Backup	1,120	0
Total	4480	2070

Table 6. Actual Sample Sizes by District

ATLANTA			
SCHOOL:	**TREATMENT**	**CONTROL**	**BACKUP**
King	202	94	
Sylvan Hills	152	74	
BOSTON			
SCHOOL:	**TREATMENT**	**CONTROL**	**BACKUP**
Cleveland	69	3	
Rogers	113	2	
Shaw	49	19	
Timilty	52	6	
Control Schools[8]		144	
COLUMBUS			
SCHOOL:	**TREATMENT**	**CONTROL**	**BACKUP**
Medina	97		
Muffin	67		
Control Schools[9]		199	
DALLAS			
SCHOOL:	**TREATMENT**	**CONTROL**	**BACKUP**
Anderson	34	54	
Edison	44	55	
Hood	36	39	
Zumwait	52	48	
HOUSTON			
SCHOOL:	**TREATMENT**	**CONTROL**	**BACKUP**
Holland	67	53	
Johnston	89	56	
MIAMI			
SCHOOL:	**TREATMENT**	**CONTROL**	**BACKUP**
Madison	47	42	
SAN FRANCISCO			
SCHOOL:	**TREATMENT**	**CONTROL**	**BACKUP**
Burbank			
Hoover	12		
Everett			
TOTAL	**1,182**	**888**	

[8] In addition to assigning students from the *READ 180* schools to the control group, Boston designated a group of 144 students from three other middle schools (Gavin, Umana and Thompson).

[9] The control group in Columbus comes entirely from three schools other than the *READ 180* schools (AIMS, Barrett and Johnson).

IV. IMPLEMENTATION

Getting schools and teachers to adopt and/or implement programs is notoriously difficult, a conclusion supported by practical experience and three decades of research about implementation. The authors of new programs for schools, would-be school reformers and analysts of school change share a common frustration: *Why do innovations—virtually all innovations—disappear in schools?* The education professional journal, *Phi Delta Kappan,* has averaged three articles a month for the last 10 years each announcing a new technique of "school reform." Yet, schools have effortlessly shrugged off all 360 reforms. Albert Shanker, the late president of the American Federation of Teachers, was fond of observing that, "Schools are the last unreformed institution from the 19th century which we are about to trundle, unchanged, into the 21st century".

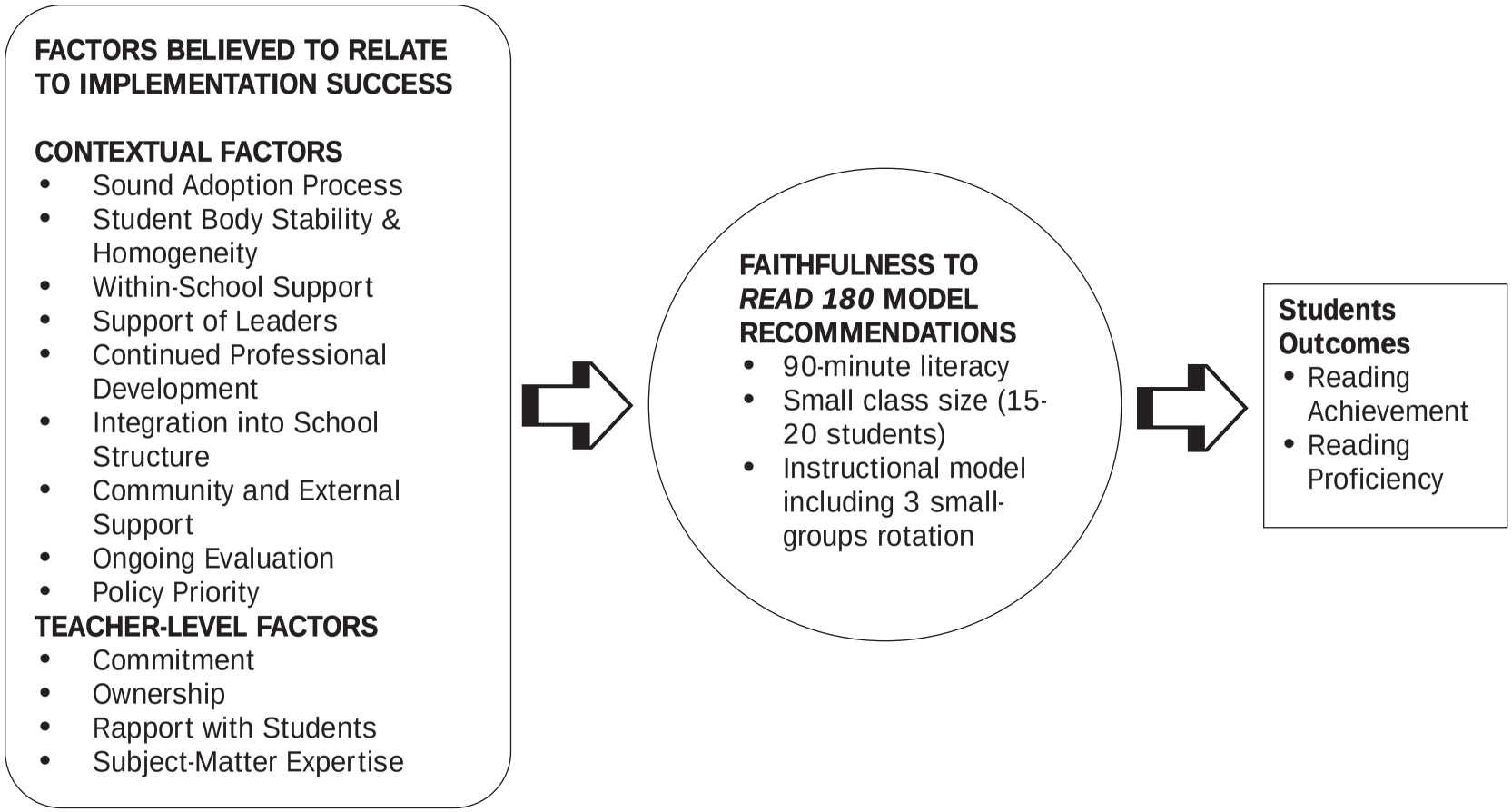
Three decades of research on school reform and improvement indicate that, whatever the efforts, the outcomes of reform are modest at best. In general, reform efforts and new programs stumble at the 'implementation' stage. That is, schools and teachers resist new programs and they are not then incorporated into the school's continuing repertoire.

The Rand Corporation's "Change Agent" critiques of the outcomes of Federal programs supporting school innovation were the first systematic documentation of the relation between the implementation process and the outcomes of school reform. Beginning in the middle 1970s, Rand concluded that Federal programs had not changed schools because schools were more adept at changing Federal programs (Mann, 1978; McLaughlin, 1990; Rand Corporation, 1974; Rand Corporation, 1975; Rand Corporation, 1975; Rand Corporation, 1975). Rather than schools eagerly accepting innovations that others had planned, schools turned the programs to their own purposes and that minimized change.

The phrase "partisan mutual adaptation" described the tension between schools and external programs. Not surprisingly, researchers find that school people believe in the correctness of what they are already doing and resent allegedly "better" programs imposed from an outside source no matter its good intentions, authoritative standing, or deep pockets. The Federal government has not transformed schools: neither have universities or foundations, nor, for the most part, state governments.

Michael Fullan (1991) has synthesized much of the implementation research and commentary in the US and Canada. "Adoption" describes the process of deciding to use a program. This is also referred to as "initiation" or "mobilization". Fullan defines initiation/adoption as "the process that leads up to and includes a decision to adopt or proceed with a change." (p 47) "Implementation" is the follow-on experience of deploying a program, adding it to the repertoire of a school or classroom. Fullan describes this as "initial use (usually the first two or three years of use)"; "the first experiences of attempting to put an idea or reform into practice." (p 48)

Fullan's work served as the basis for RAND's recent analysis of comprehensive school reform demonstration programs. In particular, RAND has documented the implementation of the New American School's Designs across the country. Incorporating Fullan's ideas and the RAND research on CSRD, the following framework depicts the complex of factors that interact and forecast the success of a new program in a school. The factors are approximations; each one may be more or less influential or important depending on the sites and the characteristics of the program.


FACTORS BELIEVED TO RELATE TO IMPLEMENTATION SUCCESS
CONTEXTUAL FACTORS
• Sound Adoption Process
• Student Body Stability & Homogeneity
• Within-School Support
• Support of Leaders
• Continued Professional Development
• Integration into School Structure
• Community and External Support
• Ongoing Evaluation
• Policy Priority
TEACHER-LEVEL FACTORS
• Commitment
• Ownership
• Rapport with Students
• Subject-Matter Expertise
FAITHFULNESS TO *READ 180* MODEL RECOMMENDATIONS
• 90-minute literacy
• Small class size (15-20 students)
• Instructional model including 3 small-groups rotation
Students Outcomes
• Reading Achievement
• Reading Proficiency

A. Contextual Factors Related to Successful Implementation

The following nine factors, adopted from Fullan and others, are generally asserted to be necessary for successful implementation of change or innovation within a school:

- Sound adoption process
- Student body stability and homogeneity
- Within-school support
- Leadership support and vision
- Integrated and continued professional development
- Community/external support
- Integration of program into school structure, curriculum, policies
- Ongoing evaluation component
- Policy priority

Throughout the remainder of this section, we consider each of these factors and our data as they relate.to *READ 180.*

1. Sound Adoption Processes

Adoption requires that relevant innovations be available to meet the needs identified, that the organization be ready to adopt new strategies, and that the resources necessary to adopt these strategies be available.

Relevance. Fullan writes about this as "Existence and Quality of Innovations". From the policy perspective, there are no problems until there are solutions. An unused solution constitutes a problem. Relevance not only means that appropriate innovations exist, but that they are also accessible to those who seek them. Fullan points out that access to innovations obviously conditions their use; for example, "resource poor" schools.

Readiness. Nominally effective schools and satisfied teachers are not fertile ground for anything new. Additionally, school leaders are a wary lot. The technology—pedagogy—on which they must rely is uncertain. Consider the perennial debates about how best to teach reading—phonics or whole language? In addition, consider the career-determining consequences of ending up on the wrong side of that chronic dispute. Unless programs are "proven" or at least have the appearance of being "evidence-based," few experienced leaders will commit to them.

Resources. Schools are plagued with too many demands and too few resources, especially time. Teaching is a performance art and a practical art. Thinking about the binary question—will teachers "teach *READ 18O"* or not—obscures a host of considerations for teachers. How will students respond? Will they be motivated? Bored? Will they act out? What will happen to classroom control? Which students will have what reaction? Will the activities for individuals and groups engage their attention? Are they feasible? Will they fit in the time allotted especially given the 'compacted curriculum'?

Consider, for example, the schools that were recruited into this test experience and how they were recruited. *READ 180* is clearly a relevant program for urban middle schools. The reading difficulties of typical urban adolescents are well-documented. Further, *READ 180* was specifically designed for middle-school students reading in the bottom quartile. In the validation districts, school officials were told to include students with low reading achievement scores, but not students reading more than 1.5 grades below grade level. Teachers reported that they had trouble identifying students who were only 1.5 grades or fewer below grade level.

Relevance notwithstanding, districts and schools specifically in the validation sites were not necessarily ready to implement a program as comprehensive as *READ 180,* nor were the resources available. The *READ 180* instructional model requires schools to find a 90-minute block of time in which to fit the program. Combined with the small class size requirement (15-20 students), *READ 180* causes major scheduling difficulties, with respect to both time schedules and personnel schedules.

Perhaps most important with respect to the particular adoption of *READ 180* is the way schools were "recruited" into the program and validation study. Scholastic initiated the program by recruiting the Council of Great City Schools. After that partnership was established, seven constituent districts were identified for inclusion in the validation study. Superintendents in each of those districts made the decision to proceed and then handed off the program to another district administrator (either a research director or a reading specialist). Those district staffers then selected schools for the project, and essentially informed building administrators and teachers that they were going to implement *READ 180.* "Ready or not, here comes *READ 180."* It was left to the schools to find the resources necessary to get the new procedures into the (new) classrooms.

2. Student Body Stability And Homogeneity

The more similar the student body in attitudes, backgrounds, and beliefs, the easier it is to implement new programs. Diversity requires flexibility and individualization, both of which — while desirable — slow and complicate implementation.

For this particular initiative, the student populations are demographically homogenous. However, student mobility rates are typically high in these districts. In some of the validation schools, new students would replace students who left the program or the school for any reason. Working new students into the *READ 180* class is difficult, more so than for a regular classroom. Due to the fairly prescriptive instructional model, *READ 180* classes are routine-oriented and students are expected to cooperate with the teacher and with each other. Further, the small class size creates a fairly intimate feeling between the teacher and the students (the adult and the adolescents). Therefore, bringing a new student up to speed on the *READ 180* routine, management suite, and in step with the rest of the "team" or "family" is complicated.

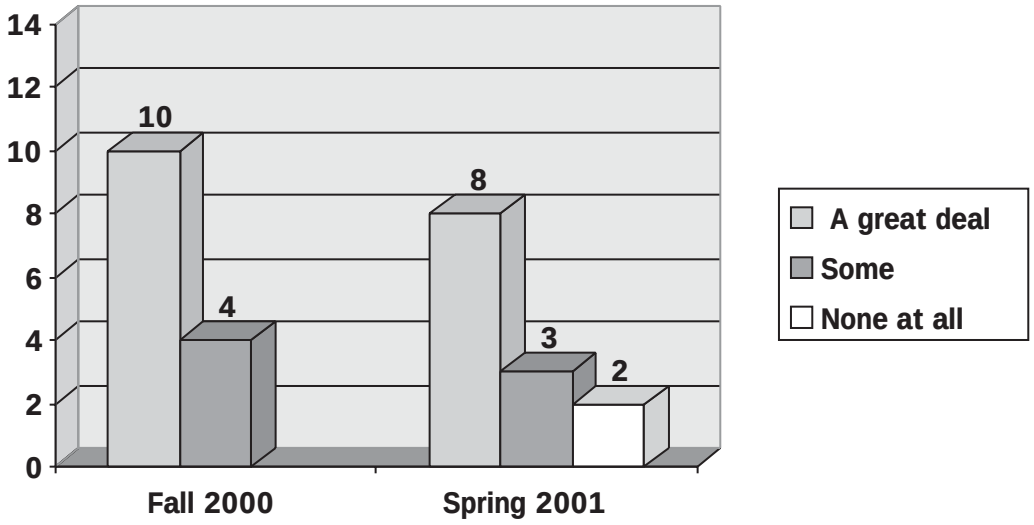
3. Administrator Support and Involvement

In discussing "Advocacy from Central Administration" Fullan says, "[i]nitiation of change never occurs without an advocate." Advocacy from the top is useful, but does not insure change. While it is undeniably helpful to have support from the central administration, schools and even individual teachers can and do initiate change and they resist and re-direct change. McLaughlin remarks that some officials adopt things for public relations purposes more than for organizational transformation purposes. In those cases, the symbolic event of "initiating a program" satisfies the appetite for improvement. The public schools of New York, for example, are legend for adopting a single pilot or demonstration version of every new idea but committing the system's one thousand school buildings to nothing.

Opposition from central administration may be more dispositive than advocacy. In general, it is easier to stop things than to start them. Many initiatives never make it to the adoption stage but are simply and effectively struck off the agenda. However, as McLaughlin ultimately concluded, "The active commitment of district leadership was essential to project success and long-run stability." (1990, p 12)

Generally, teachers in the *READ 180* validation sites report good support from building administrators, though in some cases, that support waned over the course of the year. Teachers were asked to complete a survey at the beginning of the school year and the end of the school year. One of the items on that survey asked teachers to indicate the degree of support they felt from building administrators. Figure 3 below depicts the change in feelings from the beginning to the end of the school year.

Figure 3. Teacher Reports of Change over Time in Support from Building Administration.



Whereas support from building administrators may not have been problematic in the validation sites, support from district-level administrators is a different matter. A defining characteristic of the districts where we could objectively identify consistently successful implementations of *READ 180* is support from the district office in the form of personnel. In the four "well-implemented" districts (Boston, Dallas, Houston and Columbus), at least one district official was specifically tasked as the *READ 180* liaison. These individuals may have had other roles within the district, but attending to the implementation of *READ 180* was a significant part of their portfolio. These district-level administrators organized *READ 180* meetings, helped with troubleshooting, and were generally available to support the teachers and administrators in the school buildings.

In the other three districts, where implementation of the program (and, not so coincidentally, the research design) was spotty, there was a district administrator assigned to assist with the research needs. However, this was usually an individual within the district's research or assessment department who spent little or no time insuring the implementation of the program. The teachers and administrators in the *READ 180* schools were left to fend for themselves.

4. Within-School Support and Involvement

Having teachers advocate for change is a kind of policy nirvana, desirable but hard to obtain. Fullan comments that, "the working conditions of teachers in the vast majority of schools are not conducive to sustained teacher innovations." (p 55) Since teachers are one another's most trusted sources of best practice information, strategies that maximize teacher-to-teacher contact are likely to be helpful. Examples include job-alike classroom visits, teacher-to-teacher networking, train-the-(teacher)-trainer strategies, etc.

READ 180 is no exception. The schools where *READ 180* was best implemented were those where more than one teacher taught *READ 180.* Certainly, an individual, dedicated, competent, and resourceful teacher could successfully implement *READ 180* in a single, isolated classroom. However, with a program as comprehensive and multi-faceted as *READ 180,* the better scenario is that multiple teachers have each other to lean on for support and advice. Consider, for example, the teacher who wants to run a particular student report from the *READ 180* management suite but cannot remember how to do it. Rather than having to find the time and a phone to call technical support, teachers in schools with multiple *READ 180* classes turned to each other.

Nowhere, however, is "within-school support" more relevant than with respect to the technical aspects of the program. The "Direct Instruction" rotation (where students use the *READ 180* software on the computers) is only one-third of the middle hour of the 90-minute instructional model (20 of the 90 minutes). However, when something goes wrong with the computers (a software, hardware, or networking problems), the entire model was typically disrupted.

Unfortunately, such problems occurred regularly over the course of this past academic year (2000-01), as observed during field visits and as reported on the teacher surveys. Out of the l6 teachers surveyed, 40% of the teachers reported experiencing "weekly" technical problems, while another 40% reported "occasional" technical problems.

Regular technical problems require regular technical support. Not one school that we visited had in-house technical support. The usual protocol was for a teacher experiencing technical problems to contact the district office. Typically, it would take a district technician 3-4 days to get out to the school to try to troubleshoot the problem. Then, if that technician could not fix the problem, Scholastic technical support would be contacted. Overall, a single technical problem would typically disrupt the flow of the *READ 180* instructional program for weeks at a time.

5. Community Pressure / Support / Apathy

Community pressure / support / apathy is discussed in Fullan's synthesis of others research about the problem of reform. He helpfully distinguishes a range of responses and non-responses that are possible from communities. In the presence of apathy, schools are able to do mostly what they, unilaterally, determine.

Adopting and fielding the program was the domain of the professional school people, so this implementation factor is not particularly relevant to this initiative.

6. Continued Professional Development

McLaughlin summarizes the following, professional development-related strategies as effective, “especially when applied in concert.”

- Help at the individual classroom level and from local (i.e., ‘credible’) sources
- Teacher trips to other, similar projects
- Extended, specific training for teachers
- Regular meetings focused on practical issues
- Teacher participation in project decisions
- Training for principals

Gingiss (1992) argues that to ensure implementation, it may be necessary for teachers to receive follow-up training. In the initiation stage, before teachers begin implementation, they determine the needs and benefits of the program. After implementation begins, teachers focus on mechanical tasks. Teachers then begin to develop a routine use level. At this point, it is important to receive additional support and training to maintain teacher commitment and to help teachers develop the skills necessary to refine, integrate, and renew the program.

The *READ 180* teachers involved in the validation study were first trained either in the summer or early fall just prior to the initial implementation of the program. Beyond that, districts could initiate follow-up training sessions with Scholastic, but mostly, teachers were on their own. According to the final teacher survey, two- thirds of the teachers reported that the professional development around *READ 180* was not sufficient.

Recall that *READ 180* is a comprehensive, multi-faceted program. The software component itself is sufficiently nuanced to create a need for continued training. *READ 180* is not a neatly packaged set of books that can be left at the doorstep of the classroom for a teacher to distribute to students. Nor is it simply a computer program in front of which a teacher can sit a student. In fact, a set of books and a computer program are not even half of what comprises the entire *READ 180* instructional model. Such a comprehensive program necessitates sustained support and professional development.

7. Integration into School Structure, Curriculum, Policies, Rituals, Culture and Supervision

In a system with 40 million students and 2 million classrooms, there will be a lot of variation. We count on teachers to adapt general policies to specific children and they do that. Some fraction of that 'tinkering' improves the result, some fraction degrades the outcomes.

At the level of the 'whole project', the Rand studies documented that schools and projects were in a struggle to change or to be changed. Since 1977, McLaughlin has argued that such mutual adaptation is a wholesome process. "Rand found that effective strategies promoted mutual adaptation... of a project and institutional setting to each other." (p 12) She omits to remark that schools forced more adaptation from the project to the school than *vice versa,* and second, that the school-imposed adaptation was in the direction of the *status quo ante.* In effect, schools gut projects and the net result of adaptation is no change, more of the same.

Based on our observations, we can conclude that the schools involved in the validation study forced more adaptation from the project to the school than *vice versa.* Scholastic and Interactive field researchers conducted classroom observations in the validation sites in Fall 2000 and Spring 2001. Upon reviewing the observational data, each classroom was coded as representative of one of three potential implementation models:

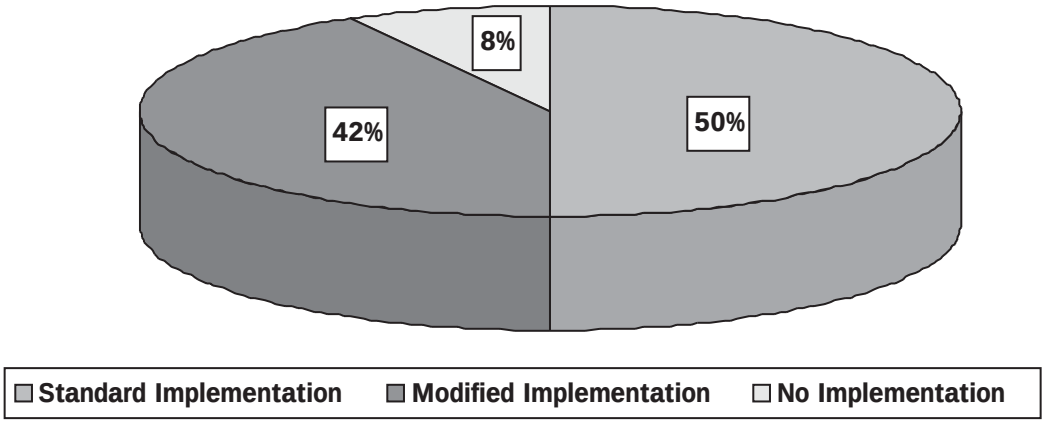
1. Standard Implementation Model
2. Modified Implementation Model
3. No Implementation [10]

Any classroom where the teachers and students completely and consistently followed the prescribed *READ 180* instructional model was classified as "standard model." Any classroom that deviated from the prescribed *READ 180* model in any significant manner (e.g. fewer than 90 minutes, no computers, etc.) was categorized as "modified implementation model." Any classroom that was intended to be a *READ 180* validation classroom, but where the program was simply not used is considered a "non-implementing" classroom.

Based on the visits by field researchers, 13 of the 26 (50%) validation site classrooms can be classified as standard *READ 180* model classrooms. Eleven of the 26 (42%) classrooms exhibit some modification to the prescribed model. Finally, *READ 180* is not being used at all in two classrooms intended to be part of the validation study.

[10] Classrooms identified as not implementing *READ 180* are excluded from further analysis.

Figure 4. Adopted Implementation Models



The table that follows displays the classification status by district and by school. We do not report specific teacher names to protect the anonymity we promised.

Table 7. Scholastic Classroom Implementation Status for Validation Districts.

DISTRICT	SCHOOL	CLASSROOM	IMPLEMENTATION STATUS
ATLANTA	King	Classroom A	Modified
		Classroom B	Standard
	Sylvan	Classroom C	Not implementing to model
		Classroom D	Modified
BOSTON	Shaw	Classroom A	Modified
	Cleveland	Classroom B	Modified
		Classroom C	Modified
	]y Rogers	Classroom D	Modified
		Classroom E	Standard
		Classroom F	Standard
		Classroom G	Standard
COLUMBUS	Medina	Classroom A	Standard
		Classroom B	Modified
	Muffin	Classroom C	Standard
		Classroom D	Standard
DALLAS	Edison	Classroom A	Standard
	Anderson	Classroom B	Standard
	Hood	Classroom C	Standard
	Zumwalt	Classroom D	Standard
HOUSTON	Holland	Classroom A	Modified
		Classroom B	Modified
	Johnston	Classroom C	Standard
		Classroom D	Standard
MIAMI	Madison	Classroom A	Modified
SAN FRANCISCO	Everett	Classroom A	Not implementing to model
	Hoover	Classroom B	Modified

8. Ongoing Evaluation

There is disagreement in the general literature about the importance of evaluation to program implementation, institutionalization and continuation. Fullan asserts that evaluations signal the importance of an event. McLaughlin found no evidence in decades of analysis of school improvement attempts that evaluations were linked to the continuation of projects, primarily because evaluations are seldom conducted.

Our experience as program evaluators suggests that teachers will frequently raise the quality of implementation they might otherwise under take simply because they are being "evaluated." In the case of this *READ 180* validation study, we have no data or reason to believe that the evaluation drove implementation in any way. In that sense, this study is a true test of the impact of *READ 180* under world-of-practice conditions.

9. Policy Priority

Every topic must compete with state standards-based instruction and that instruction is centered on basic skills. Reading, language arts and mathematics are more than rhetorical priorities—their instruction is supervised, monitored and enforced in many instances with career consequences for school employees. In Chicago, for example, basic skills instruction is literally scripted. Everything else is secondary and potentially expendable. Further, the expectation is that high priority, official programs will be more faithfully implemented than others will. Mandates trump extracurriculars.

READ 180's status as a literacy program targeting the neediest students would generally give it high policy priority. And, in most cases, school administrators and teachers went out of their way to get *READ 180* implemented, even if its structure needed to be modified. However, in one district, we saw *READ 180* bumped from the top of the policy priority totem pole. In Atlanta, the district adopted *Success for All* in all the middle schools in the district. The first 90 minutes of every instructional day was spent teaching literacy in the *Success for All* mode. Those students who were designated for *READ 180* would follow the 90 minutes of SFA reading instruction with 70 minutes (a model modification) of *READ 180.* When there was any conflict or a need for extra time for other parts of the curriculum, it was *READ 180* fidelity that was sacrificed.

B. Teacher and Classroom Factors Related to Successful Implementation

Rand's "Change Agent" studies provided evidence to elevate classroom teachers to the single most influential determinants of education policy. Regardless of Federal programs, State mandates or local priorities, teachers eventually take their children into the classrooms, close the doors and decide which children get what quality of instruction in which topics. The following factors make classrooms relatively impervious to most outside influences:

- The press of the teachers' core business—the daily care of 20-plus children in a class (or a hundred children in a middle school specialized teaching assignment)
- The logistic impossibility of directly supervising 2.2 million elementary and secondary classroom teachers
- The belief that decisions about teaching and learning are best left to the people closest to the children—teachers
- The political protection afforded by their unions (school teachers are the most unanimously organized sector of the American economy).

This is a nuanced argument. Most of the implementation literature concludes that without cooperation from teachers, nothing happens. Teacher cooperation, however, does not guarantee change. Teachers do not sign purchase orders. In addition, they do not control the general outlines of the topics that are to be taught. Just as the classroom is loosely coupled to other levels of governance, so also are teachers loosely coupled to each other.

Therefore, in addition to the contextual factors described above, program implementation is thought to be a function of the following four teacher-level factors:

- Commitment to the program
- Level of ownership
- Rapport with students
- Subject matter expertise

1. Commitment to the Program

Most research indicates that teacher commitment is a necessary condition for the successful implementation of school innovations. Teacher "buy-in" maximizes the prospects for change (Berend, 2000). For instance, Cooper (1998) found that teacher resistance to *Success for All* (a school reform curriculum) was the biggest prohibition on quality implementation.

Our data, both from observations and teacher surveys, suggest that almost all of the teachers involved in the validation study were committed to the program. As noted above, frequently, the program was adapted to the school rather than the other way around. Yet, teachers showed strong commitment to the basic premise of the program. Most telling is the fact that 93% of the teachers surveyed stated that they would recommend *READ 180* to reading teachers at other middle schools.

2. Level of Ownership

Datnow and Castellano (2000) identified the following themes as indicators of teacher-level ownership:

- Teachers made adaptations to the curriculum
- Teachers implemented the program despite believing that it constrained teacher autonomy and creativity because they thought it benefited students
- Teachers were skeptical about the program because it was developed by an outside entity (negatively related to ownership)

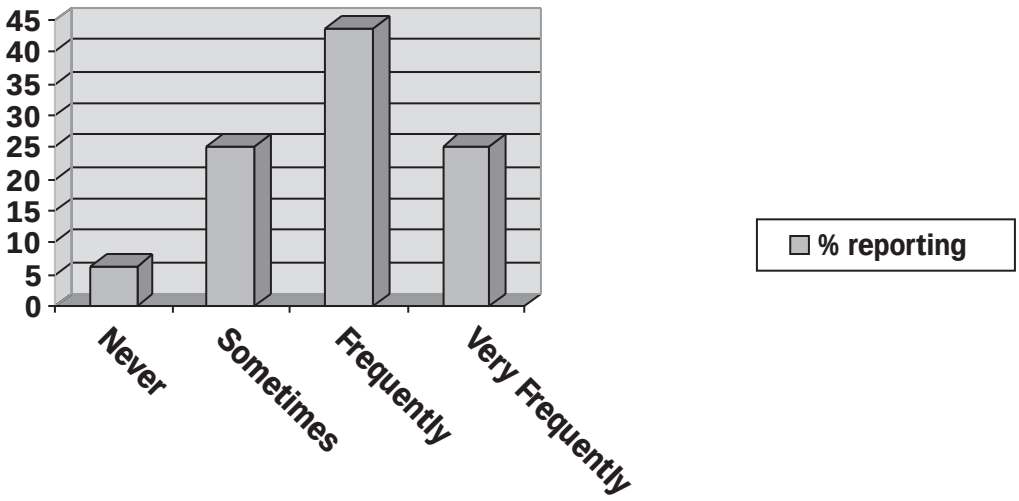
"Variation in curriculum implementation is inevitable." (Datnow and Castellano, 2000, p. 779) Flexible school policies, the lack of accountability, or variation in teacher perceptions of what works for students are all sources of implementation flutter. In general, teachers are skeptical that 'outsiders' can 'really know (their) children'. That skepticism is a barrier to acceptance and ownership.

Once again, in half the classrooms in the study, the program was adapted to the school. However, even within classrooms, we found that the majority of teachers made adaptations to *READ 180.* For example, in one school, teachers and students set aside 45 minutes of the 90-minute block twice a week to focus on a writing skills curriculum.

There are additional or optional elements of the *READ 180* program that teachers might integrate into the routine. The "Quick Writes" program allows teacher to add a writing component to the *READ 180* instructional model. In addition, the "Reading Counts" quiz program is bundled with the *READ 180* software and allows students to be quizzed each time they complete a paperback during the independent reading rotation. Finally, the "Book Expert" is a part of the management suite that allows teachers to print up a list of books that match students' self-reported interests and reading ability level.

Of the 16 teachers who completed and returned a survey in the Spring, 10 told us that they had students complete a "Reading Counts" quiz every time they completed a paperback or audiobook. The other six utilized "Reading Counts" "occasionally." Similarly, 10 teachers reported ever using the "Book Expert." The figure below shows the breakdown of the frequency of use of the "Quick Writes" program by teachers and students.

Figure 5. Frequency of Use of the "Quick Writes" Program



3. Rapport with Students

The relationship teachers have with students may not predict the fidelity of implementation, but it may affect student outcomes. If students are more willing to credit the instruction of teachers they respect and who understand them, they may more fully trust the teacher's message; this may enhance the effectiveness of a program.

4. Subject Matter Expertise

In general, teacher subject expertise is related to student learning. In the research literature, certification or licensure is a proxy for subject matter expertise and it is positively related to student learning. (Darling-Hammond et al., 1990 and Goldhaber & Brewer, 2000).

V. STUDENT OUTCOMES

A. Student Reading Achievement

1. Special Considerations: Measuring Growth in Achievement.

It is deceptively simple to look at student achievement data at two points in time and see what change, if any, has occurred. However, remarking change and understanding the change are two very different notions. In fact, the interpretation of change is fairly complicated. Three problems are discussed below.

(a) What Constitutes "Normal" Growth?

Any form of test scores can define growth, but the most commonly used are grade equivalents (GE), percentile ranks (PR), scaled scores (SS), and normal curve equivalents (NCE). The GE scale is constructed so that, for average students, there is a 1.0 increment from grade to grade. So, for example, the average 4th grade student will score at 4.0 at the beginning of Grade 4. Thus, there is some ease of interpretation with grade equivalents. However, the unequal units of the GE scale result in different growth rates for students scoring at different points on the scale. Therefore, it is a very poor scale for measuring growth.

When Percentile Ranks are used, "normal" growth is defined as maintaining one's position relative to the norm group. The 4th Grade student who tested at the 9Oth percentile one year and again at the same time the next year has exhibited normal growth. The problem with using PR for measuring growth is that percentile ranks do not represent an equal-interval scale. A 15-point change on part of the scale is not the same as a similar change on another part of the scale.

To solve the problem of inequality of units on the PR scale, Normal Curve Equivalents (NCE) are derived. NCEs result from dividing the normal curve into 99 equal units. For this reason, NCEs have become the preferred mode for measuring change.

An equally preferable method of measuring change is with Scaled Scores (SS). Normal growth in scaled scores occurs when students increase by as many units as the standardization group. SS offer the advantage of measuring growth on a continuous scale across levels and forms of the test. They should NOT, however, be used to make comparisons ACROSS subtests.

If an experimental program is being evaluated, a control group should be selected from the same population and both groups should be pre- and posttested. The pre/post difference ("gain") for the control group would define "normal" growth. The performance of the experimental group could then be compared to that baseline. Our original research design called for exactly this pattern, and we have held to that design where we have the data.

(b) Regression toward the Mean.

A general principle of measurement is that students who score very low on a test tend to score higher on a retest, and students who initially score very high, score lower on the retest. Students scoring at the extremes tend to regress toward the mean of the total group. So, while some gains or losses may be actual or real, some changes are due to this statistical artifact of regression toward the means.

(c) The Unreliability of "Gain" Scores.

While it is hoped that student performance is consistent ("reliable") from one test administration to another, pre/post change scores are less reliable than either of the scores from which the growth score is derived.

The best ways to increase the reliability of gain scores are to keep the time between pretesting and posttesting as long as possible, and to administer more than one posttest. Thus, the "gain" score results for an evaluation are less reliable after one year than after two.

2. Standardized Achievement Tests

(a) Tests Administered.

As mentioned earlier, the original research design called for the validation districts to administer the following two parts of the Stanford Achievement Test Series (Ninth Edition) ("*Stanford-9*"):

- Multiple-choice Reading subtests (reading vocabulary and reading comprehension). These subtests are also aggregated into a single Total Reading score
- Open-Ended Reading Assessment

These measures of student reading achievement were to be administered to all treatment, control and backup students just prior to the implementation of *READ 180*, and at the end of the academic year. However, as with the sampling framework, despite the fact that district administrators had agreed to this assessment protocol, the reality was different from the original plan.

Table 8. Actual Standardized Test Administration Schedule

DISTRICT	PRETEST		POSTTEST	
	TEST	DATE	TEST	DATE
ATLANTA*	*Stanford-9* *Multiple-choice reading	Spring 2000	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2001
BOSTON	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2000	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2001
COLUMBUS	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2000	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2001
DALLAS	*Stanford-9* *Multiple-choice reading	Spring 2000	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2001
HOUSTON	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2000	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2001
MIAMI	*Stanford-9* *Multiple-choice reading	Spring 2000	*Stanford-9* *Multiple-choice reading	Spring 2001
SAN FRANCISCO	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2000	*Stanford-9* *Multiple-choice reading *Open-ended reading	Spring 2001

(b) District-by-District Reading Achievement Results.

Following are district-by-district graphical depictions of pre- to posttest growth in reading achievement. Due to variability in grade levels experiencing *READ 180* (see table above), analysis were conducted separately by district. Further, only districts providing tests scores (Boston, Dallas and Houston) could be included in the present analysis.

A few main findings emerge from the data depicted below.

- In Boston (6th grade students) and Dallas (8th grade students), the differences in growth between the treatment (*READ 180*) and control groups are statistically significant and in favor of the *READ 180* students.
- The treatment and control groups in Boston are non-equivalent groups—the students in the *READ 180* classes achieved lower in the Spring of 2000 than the control group students. After a year in the *READ 180* program, however, the treatment group students have caught up to the control group students.
- In Houston, the reverse situation is true. The control group students start out lower than the *READ 180* group. While the *READ 180* students make similar, or perhaps even larger gains than *READ 180* students in the other district, those gains are not significantly different than the control group students.

Figure 6. Boston (6th Grade)

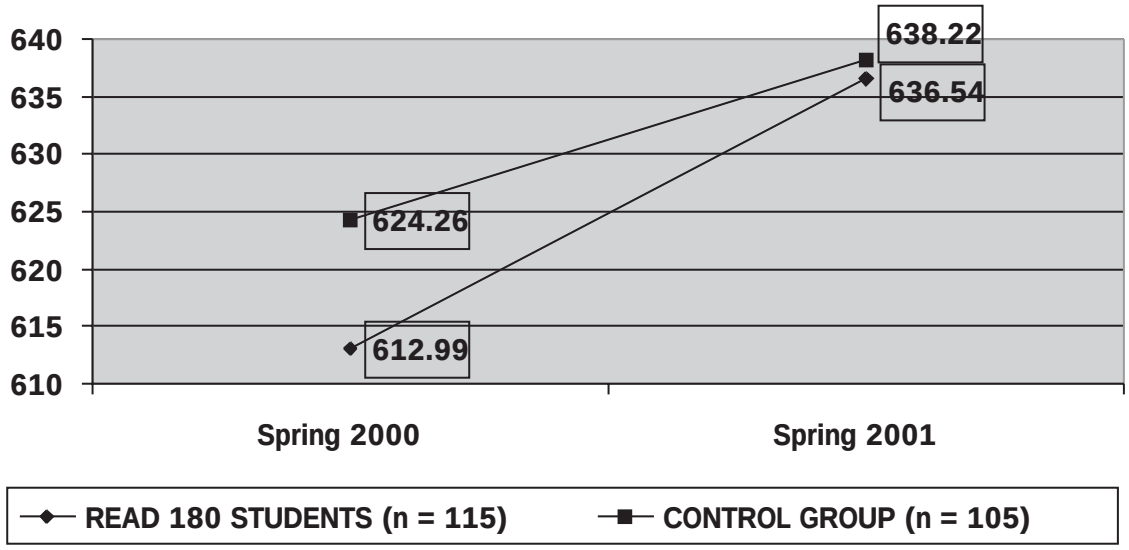


Figure 7. Dallas (8th Grade)

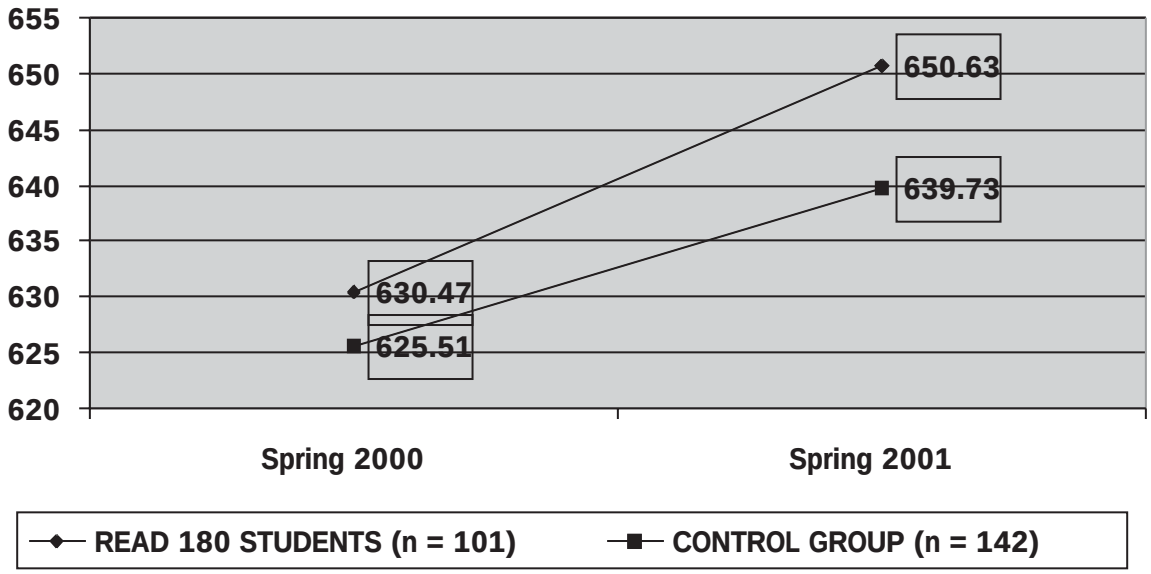


Figure 8. Houston (7th Grade)

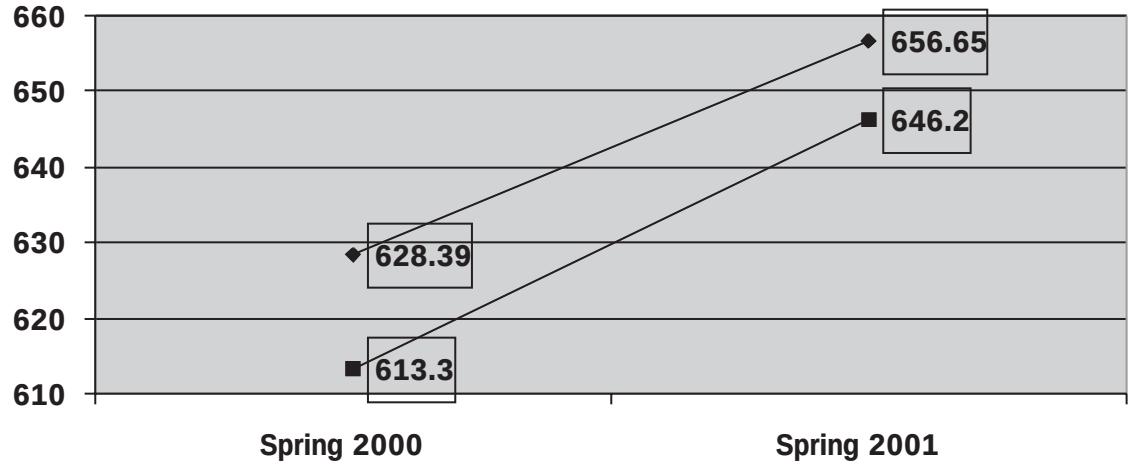
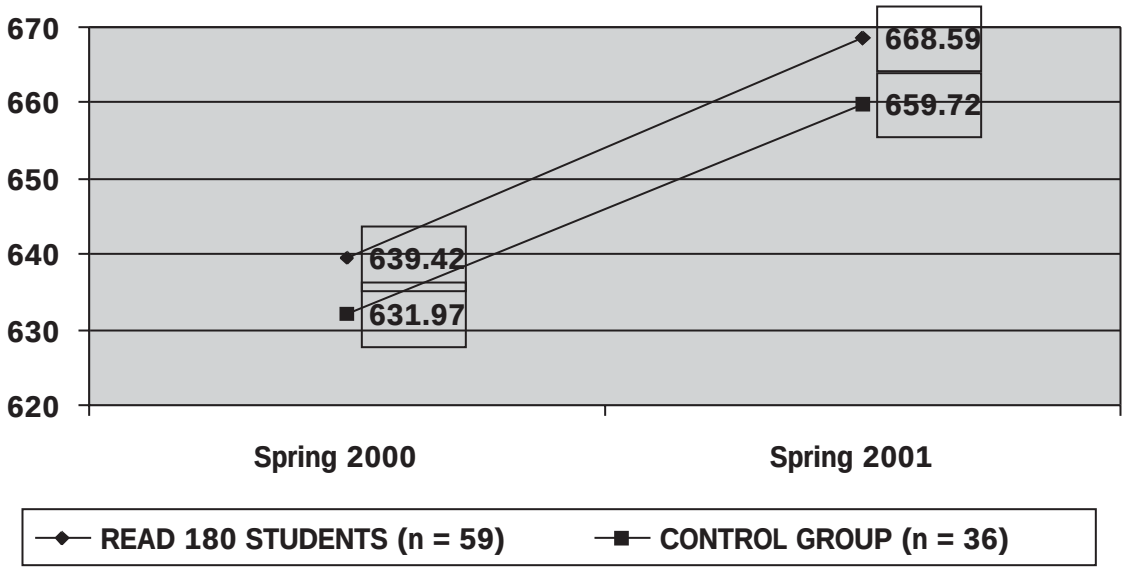


Figure 9. Houston (8th Grade)

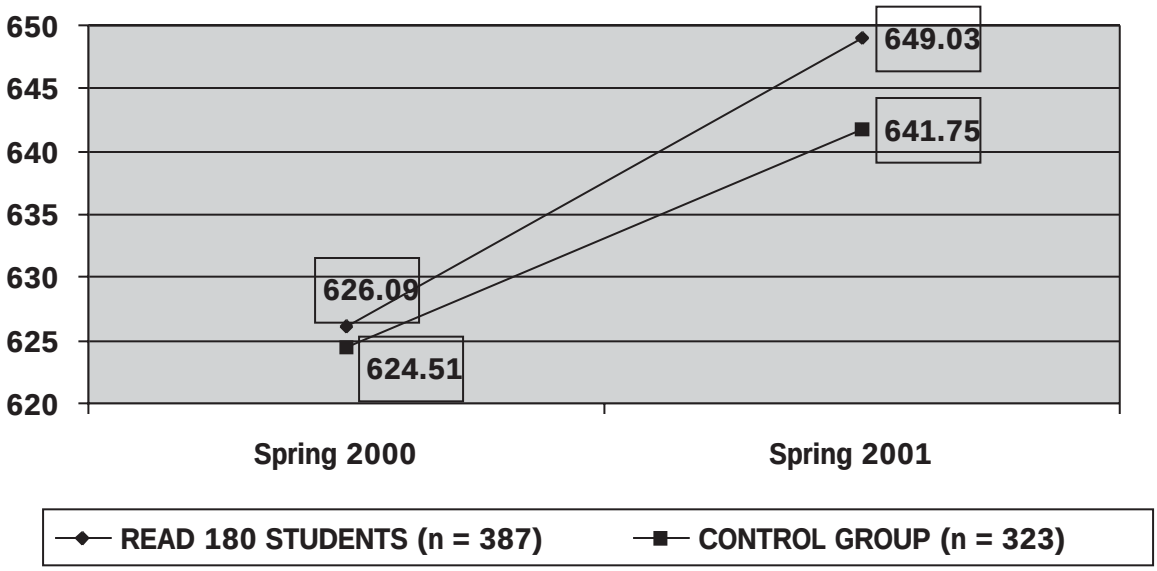


(c) Combined Achievement Growth.

The Stanford scale scores express performance on all forms of a given subtest along a single scale. Once a raw score has been converted to its scaled score with the Stanford system, you no longer need to be concerned with the level or form that was administered. This makes scaled scores especially suitable for comparisons when different forms or levels of the battery have been administered and for studying change in performance over time. Scaled scores have the advantage of representing approximately equal units on a continuous scale; that is, a difference of 5 points between two students' scores represents the same amount of difference in performance wherever it occurs on the scale.

Therefore, given these psychometric properties of Stanford-9 scale scores, we can combine students in all three districts across all three grades. The following figure shows the growth in achievement for 387 *READ 180* students compared to 323 control students. **The difference in the growth between the treatment (+22.94) and control groups (+17.24) is statistically significant (F12.624, α = .000) and in favor of the students in the classes.**

Figure 10. All Students (3 Districts, 3 Grade Levels)



3. Analysis of Covariance

In most social science research, and especially educational research, individual differences account for the bulk of the variance of phenomena studied. So, for example, in a study such as this one, where student achievement is the outcome, it is safe to assume that the variance (range, spread, etc.) in achievement scores is mostly due to differences between individual students. In that vein, there are two possible analytic approaches: treat individual differences as error or focus the study on those differences.

The former approach originates in experimental psychology where the goal is to control behavior and variation within treatments; individual variation becomes "error variance." Those trained in the experimental approach rely, almost exclusively, on analysis of variance. The analysis in the section immediately prior to this one are consistent with the "quasi"-experimental approach. The latter approach comes from correlational psychologists who regard individual and group variations as important effects of biological and social causes. Those trained according to this discipline examine data using correlation and regression analysis.

Fortunately, these approaches are not mutually exclusive. There are analytic techniques available that allow one to study group differences while taking individual differences into account. In particular, analysis of covariance (ANCOVA) is an analytic procedure available for considering individual differences without having to give up the notion of differences among groups, settings, and/or treatments.

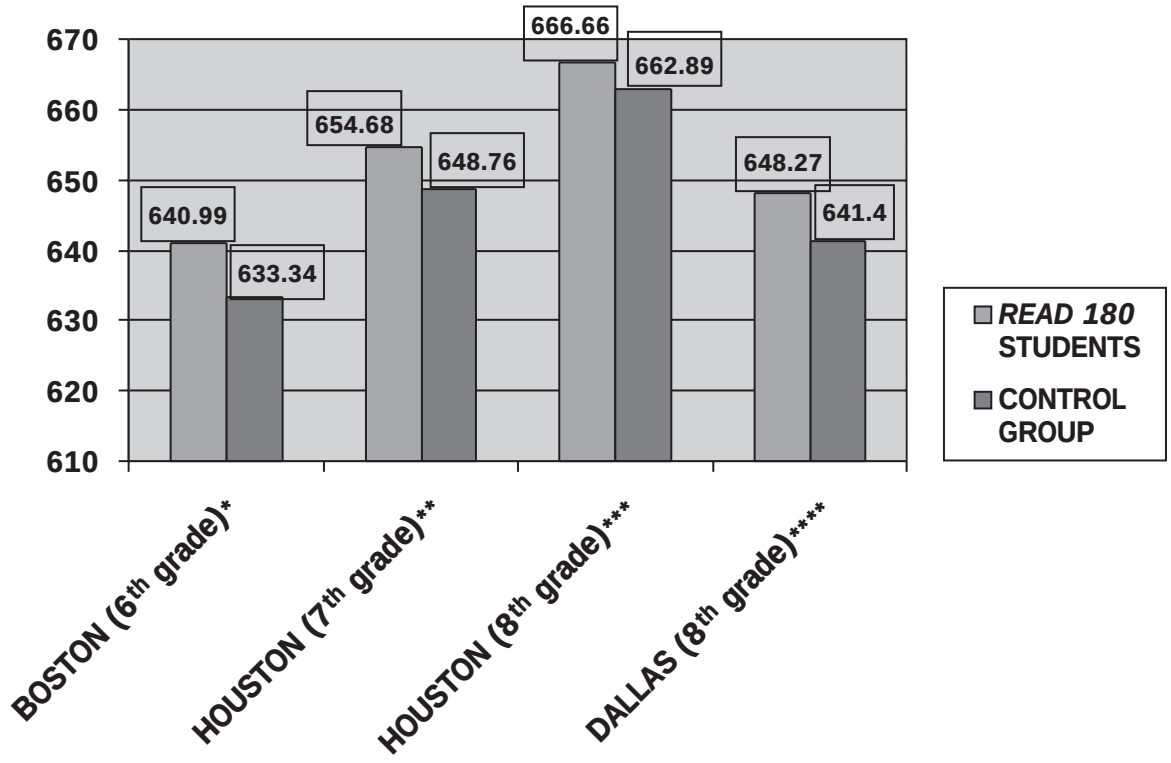
In common language, ANCOVA allows us to examine differences on a dependent variable of interest (student achievement) between two or more groups (treatment and control), "controlling for" or "accounting for" a relevant attribute (e.g. prior academic performance) measured prior to assignment to groups. In our case, Spring 2001 Stanford-9 scores are the dependent variable of interest, and we are controlling for performance on the Stanford-9 administered in the Spring of 2000, which serves as a measure of difference prior to the intervention.

The figures that follow graphically depict the results of our Analysis of Covariance. Very generally, the graphs show, for each grade level, the mean score which has been adjusted based on the covariate (the measure of prior ability), the Spring 2000 Stanford-9. Stated another way, we could have simply compared the performance of the three groups on the Spring 2001 proficiency test. But, since we recognize that individual students within the groups start at different achievement levels, ANCOVA was used to generate mean scores that are "adjusted" based on prior performance as measured by the Stanford-9. Using ANCOVA has allowed us to look at group differences while taking individual differences into account.

The findings that emerge from the analysis of covariance confirm what we learned from the previous growth analysis:

- In Boston (6th grade students) and Dallas (8th grade students), the differences on the adjusted mean score between the *READ 180* groups and the control groups are statistically significant (Boston–F=9.614. α= .002; Dallas – F=7.188, α= .008) and in favor of the *READ 180* students.
- There are no significant differences in Houston, perhaps mostly because of the low number of students involved in the analysis.
- The partial eta^2 is a statistic that estimates the magnitude of the effect. It estimates the proportion of the variance in the dependent variable that can be explained by knowing only the group to which a student belongs. The partial eta^2 in Boston is .042, meaning 4.2% of the variance in the Spring 2001 scores is a function of group differences; i.e. student assignment to the *READ 180* program.

Figure 11. Spring 2001 Stanford-9 Total Reading Adjusted Means

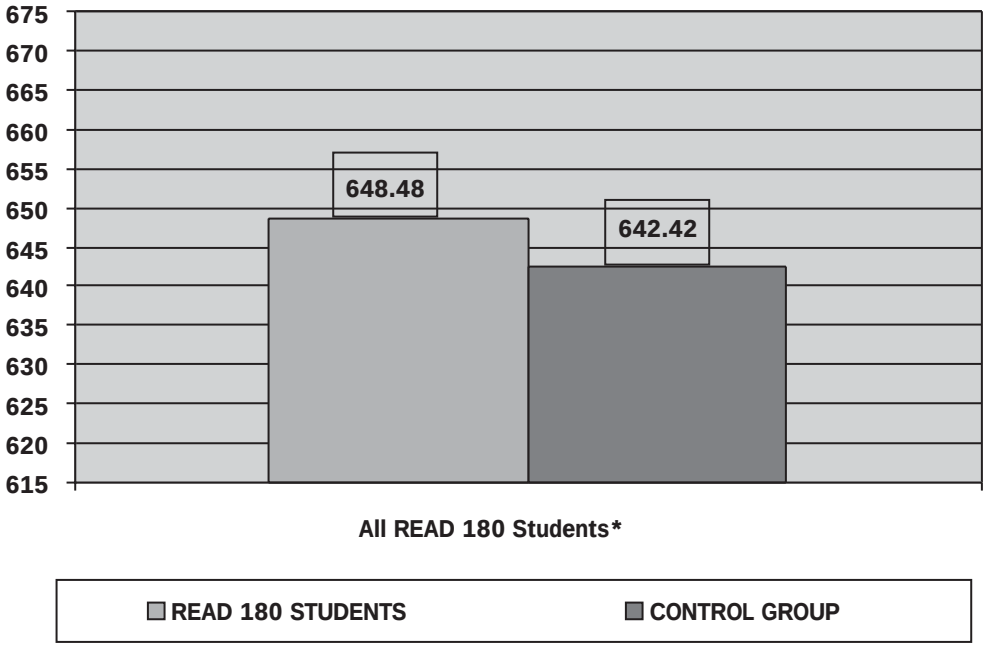


*Evaluated at covariates appeared in the model: Stanford-9 (Spring 2000)—TOTAL READING = 618.37 (Partial eta^2 = .042)
**Evaluated at covariates appeared in the model: Stanford-9 (Spring 2000)—TOTAL READING = 627.57 (Partial eta^2 = .029)
***Evaluated at covariates appeared in the model: Stanford-9 (Spring 2000)—TOTAL READING = 621.83 (Partial eta^2 = .014)
****Evaluated at covariates appeared in the model: Stanford-9 (Spring 2000)—TOTAL READING = 636.60 (Partial eta^2 = .007)

As with the growth analysis, since we are using scale scores, we have combined students across all three districts and all three grade levels for analysis of covariance.

Once again, **the difference in the adjusted mean between the treatment (648.48) and control groups (642.42) is statistically significant (F=15.396, α = .000) and in favor of the students in the classes.** The partial eta ^{2} is .021.

Figure 12. Spring 2001 Stanford-9 Total Reading Adjusted Means



*Evaluated at covariates appeared in the model: Stanford-9 (Spring 2000)—TOTAL READING = 625.37 (Partial eta^2 = .021)

4. Columbus Public Schools

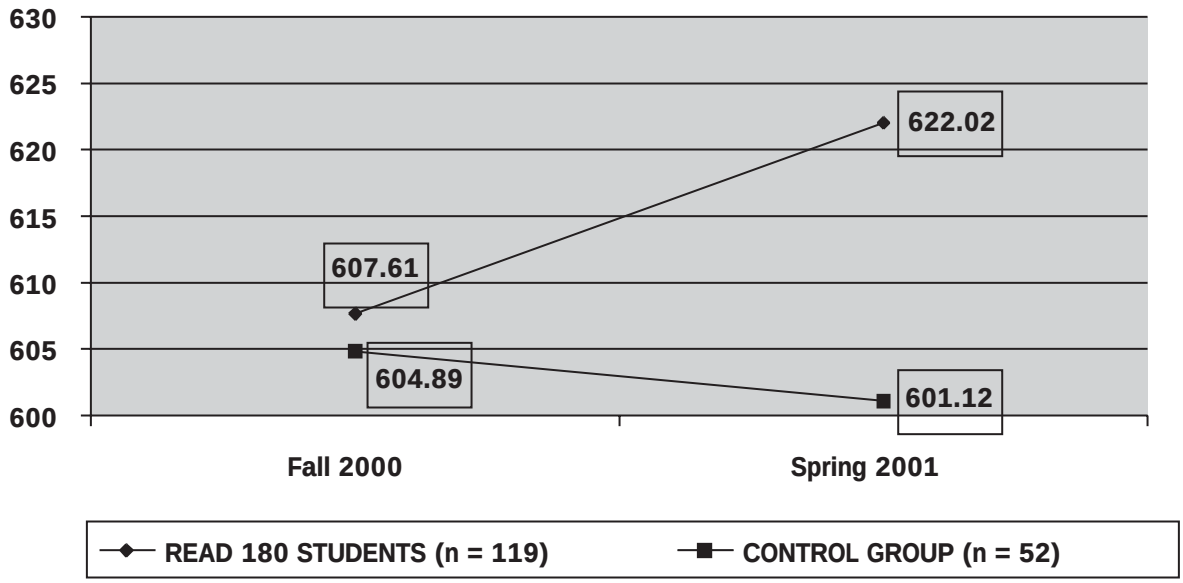
READ 180 was implemented reasonably well in 6th and 7th grade classrooms in the Columbus, OH Public Schools. And while the district's instructional and assessment protocols prevented them from fully cooperating with the research design, district officials were extremely cooperative and supportive of the idea of collecting any data necessary to conduct a separate evaluation of *READ 180* in Columbus.

For this evaluation, Columbus Public Schools administered the Stanford-9 Reading Comprehension subtest[11] to students prior to implementing *READ 180* (Fall 2000) and again at the end of the academic year (Spring 2001).

Figure 13 below compares the progress of the students in the *READ 180* classes to a control group over the course of this past academic year. The *READ 180* students made significantly greater progress (+14.41 scaled score points) than the control group (-3.67).

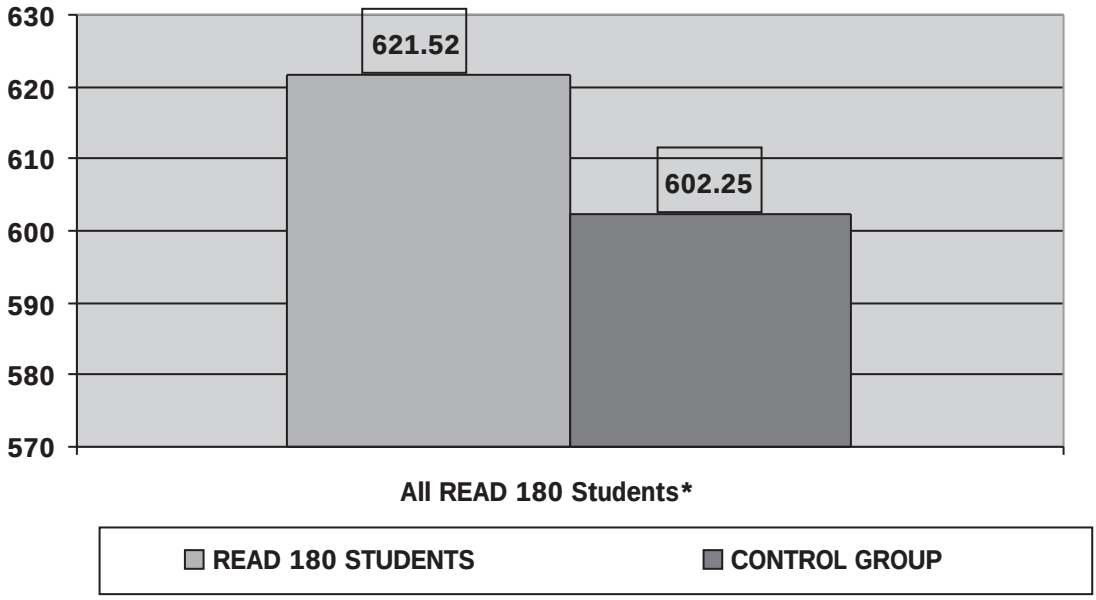
[11] Reading comprehension is one of two subtests that comprise the Total Reading score reported elsewhere in this report. For an unspecified reason, only 4 students took the Reading Vocabulary subtest (the other subtest of the Total Reading score) in the Fall and again in the Spring. Therefore, we do not report Reading Vocabulary or Total Reading scores for Columbus.

Figure 13. Reading Comprehension



Using analysis of covariance, which yields adjusted means for the Spring 2001 scores controlling for performance on the Fall test, we see again that students exposed to *READ 180* significantly (statistically and practically) outperformed the control group students in reading comprehension.

Figure 14. Adjusted Means for Reading Comprehension



*Evaluated at covariates appeared in the model: Stanford-9 (Fall 2000)—READING COMPREHENSION = 6606.78 (Partial eta^2 = .091)

B. Student Reading Proficiency

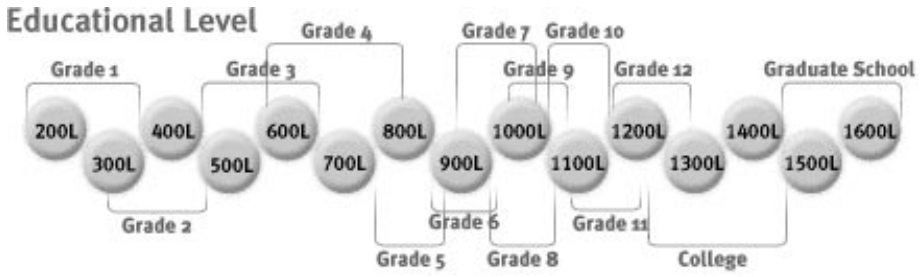
1. The Scholastic Reading Inventory (SRI) and the Lexile Framework.

The Scholastic Reading Inventory consists of a series of multiple-choice tests designed to measure reading comprehension. SRI scores are determined by the Lexile Framework® and provide direction for matching students with appropriate reading material.[12]

The Lexile Framework® is a tool for looking at a reader's ability in relation to the difficulty of specific texts. MetaMetrics, Inc., an educational measurement and technology firm, developed the Lexile Framework® over a 10-year period with funding from the National Institute of Child Health Development (NICHD). The same scale is used for both readers and text material, providing an empirical method for predicting the match between the reader and text. The difference between a reader's Lexile Measure and a text's Lexile Measure is used to forecast how well the reader will comprehend the text.

The Lexile Framework® measures the difficulty of text using transformations of sentence length and frequency of word usage. The raw score performances of readers are converted into Lexile Measures (L). When readers are well-targeted (i.e., the difference between text and reader is near OL), the reader's comprehension is about 75%. When a text measure exceeds a reader measure by 250L, the reader's comprehension drops to about 50%. When a reader measure exceeds a text measure by 250L, the reader's comprehension is about 90%.

Figure 15. Lexile Framework®

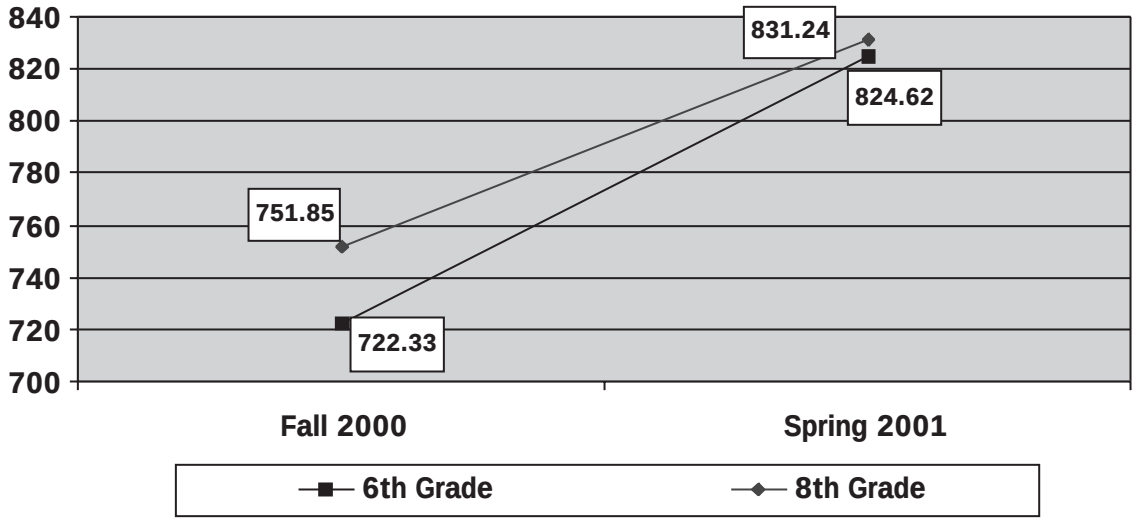


[12] The most recent Scholastic Reading Inventory -- Interactive Version measures students' comprehension against norms established by nationwide samples and is computer ready. The results give teachers and administrators a snapshot of a student's reading level—one that corresponds to widely used standardized tests.

2. Student Reading Proficiency as Measured by the SRI

As part of the research design (and probably as good instructional practice within the *READ 180* program), we asked all the *READ 180* teachers to administer the SRI three times over the course of the academic year. Not all teachers complied with the request. Therefore, the 63 6th grade students included in the analysis depicted below are from Boston, and the 72 8th graders are from Dallas and Houston [13] combined. As depicted in Figure 14, students in the *READ 180* classes made sizable gains on the SRI from the beginning of the year to the end of the year.

Figure 16. Scholastic Reading Inventory (SRI) Pre- and Posttest Scores



According to MetaMetrics, Inc., the company that developed the Lexile framework, the average middle school student grows about one to two lexiles per week. The average primary school student grows about three to four lexiles per week. Therefore, over a full, 36-week academic year, middle school students should expect to grow 90 to 140 lexiles. In Boston (the 6th graders above), the average student grew by over 102 lexiles over the course of the school year. In Dallas and Houston (the 8th graders above), the average student gained almost 80 lexiles. Recall, that students were selected for this program because they were at least 1.5 grade levels behind. The reported growth is for students who are typically labeled "non-readers." Considering that the SRI was probably administered closer to 25-30 weeks apart, the growth in reading proficiency exhibited by the students in the *READ 180* program is very positive and much greater than would otherwise have been the case.

[13] We had pre- and posttest data from seven 7th grade students in Houston. We did not report those results because of the small number of students.

VI. IMPLEMENTATION VS. OUTCOMES

The foregoing sections indicate some benefits to student reading achievement and student reading proficiency from *READ 180*. Those findings from the treatment group represent an aggregate of students who have been in classrooms with varying levels of implementation fidelity, thus masking the effects of high-fidelity implementation.

Our aspiration was to provide fair tests of *READ 180* through inclusion of implementation data.[14] A key use of implementation data is to divide the treatment sample into high- versus low-fidelity classrooms, expecting that any treatment effects are to be found under conditions of the highest fidelity (i.e., the foregoing achievement test results may represent a masking of true findings beneath variability in implementation fidelity).

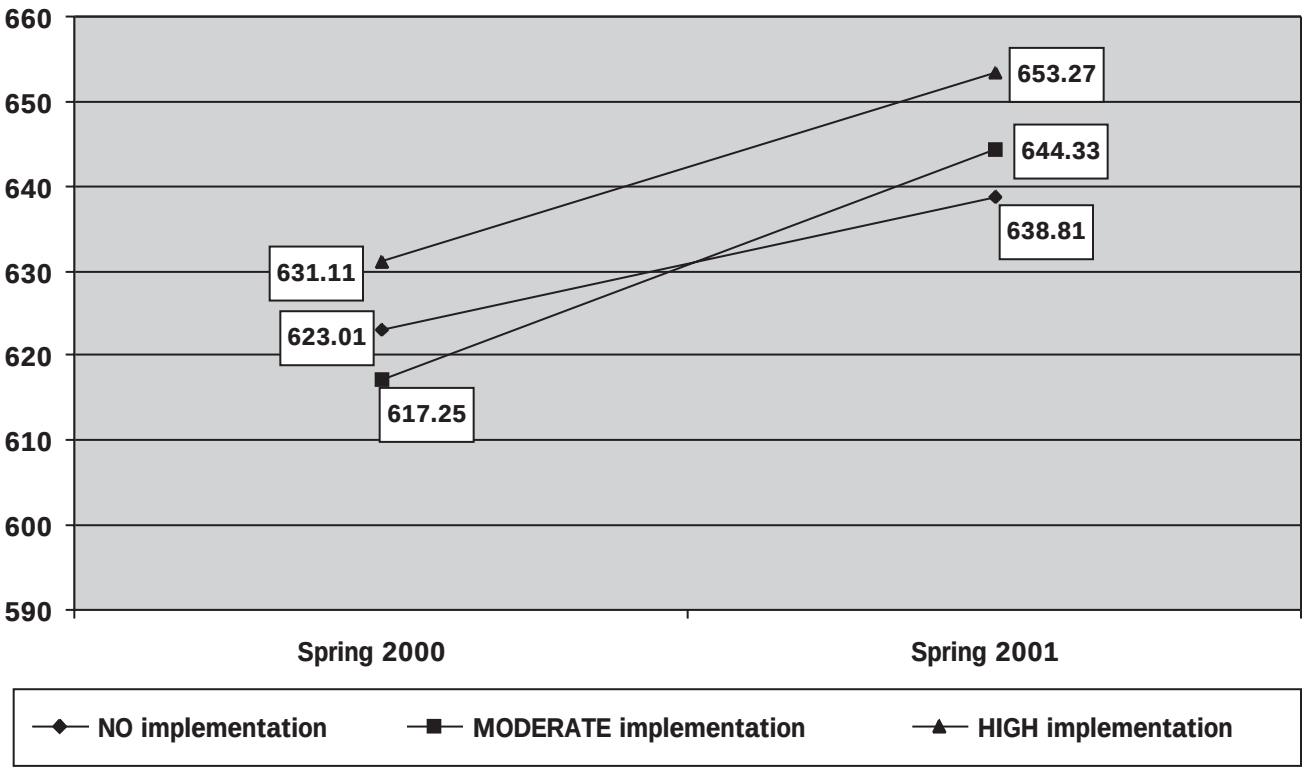
Using survey data and observational data, classrooms were coded based on the quality of implementation. Factor analysis were performed on the teacher survey data, and factor scores were compared to observational data from the classrooms. Using the factor scores and the observational data, *READ 180* classrooms were categorized as either "moderate" or "high" quality implementation. The control groups were then added as representative of classrooms with no implementation.

NOTE: The analysis of test scores against level of implementation are not as robust or reliable as we would have liked. Of the 8 classrooms coded as representative of "high-quality" implementation, 4 are from Dallas (8th grade students). In fact, all of the validation classrooms in Dallas were rated as high-quality *READ 180* classrooms. Since Dallas used *READ 180* with 8th grade students, the result is non-equivalent groups.

[14] Unfortunately, we expected that the data exported from the student management suites would be our best source of implementation data. The management suite maintains extremely fine-grained data about "use," including the total number of minutes students access the system, the total number of words read, etc. Such data could be used in the same way "dosage" data are used in pharmaceutical research. However, those data turned out to be entirely unreliable. Mostly, technical difficulties caused the loss of data in many cases. Some teachers believed the system was not tracking usage properly, and others pointed out the inaccuracies due to data loss. We tried to maintain some integrity in those data, but that proved impossible in the end.

However, the figure below shows that the group of students with no exposure to *READ 180* did not make the same kind of academic progress as either of the two treatment groups.

Figure 17. Academic Progress



VII. CONCLUSIONS

A. IMPLEMENTATION

- *READ 180* is a comprehensive, multi-faceted program. *READ 180* is not a neatly packaged set of books that can be left at the doorstep of the classroom for a teacher to distribute to students. Nor is it simply a computer program in front of which a teacher can sit a student without any monitoring. In fact, a set of books and a computer program are not even half of what comprises the entire *READ 180* instructional model. Such a comprehensive program requires sustained support (both internal and external) and professional development. The software component itself is sufficiently nuanced to create a need for phased training.
- Leadership counts. A defining characteristic of the districts where we could objectively identify consistently successful implementations of *READ 180* is support from the district office in the form of personnel. In the four "well-implemented" districts (Boston, Dallas, Houston and Columbus), at least one district official was specifically tasked as a *READ 180* liaison. Attending to the implementation of *READ 180* was a significant part of their overall responsibilities in the district. In the other three districts, where implementation of the program was spotty, the district administrator assigned spent little or no time tending to the implementation of the program. The teachers and administrators in the *READ 180* schools were left to fend for themselves.
- Hardware and software problems frequently disabled the instructional reading component of the program in many sites, and technical support was not always available to assist teachers in a timely manner.
- Of all the classrooms intended to be part of the validation study, only half implemented *READ 180* according to the standard design (42% implemented a modified model).
- Teachers report that the Scholastic management suite is helpful for addressing student needs; however, the majority of teachers are not accessing its full potential.

B. OUTCOMES

- Across all *READ 180* students in the three districts that provided us with test scores (Boston, Dallas and Houston), the difference in the growth on the Stanford-9 between the treatment (+22.94) and control groups (+17.24) is statistically significant (F=1 2.624, α .000) and in favor of the students in the *READ 180* classes.
- In Columbus, where pretests were administered in the fall of academic year 2000-01, the difference in the growth on the reading comprehension subtest of the Stanford-9 between the treatment (+14.41) and control groups (-3.77) is statistically significant (F=13.182=.000) and in favor of the students in the *READ 180* classes. The differences in growth in Columbus rise to the level of considerable educational significance.
- Using analysis of covariance to control for prior levels of achievement, the difference in the adjusted mean between the treatment (648.48) and control groups (642.42) is statistically significant (F=12.624, α =.000) and in favor of the students in the *READ 180* classes.
- Control group students, i.e. students with no exposure to *READ 180*, did not exhibit the same kind of positive academic progress demonstrated by students in classrooms where implementation quality was classified as either moderate or high.
- The growth in reading proficiency exhibited by the students in the *READ 180* program is very positive (80-100 lexiles) and greater than expected over the treatment period.

APPENDICES

APPENDIX A INTERACTIVE, INC. CAPABILITIES

Interactive, Inc. is a leading program evaluation firm that specializes in the empirical documentation of the impact of learning technologies and in evaluating school and classroom improvement programs. Interactive's research has been featured in *American School Board Journal, Electronic School, eSchool News, The Educational Administrator, Education Leadership* and *Phi Delta Kappan*. The Organization has 30 years of experience working with schools and companies in ways that minimize demands and maximize validity, reliability, practical significance and public visibility. A sample of our recent and current research includes:

1. For the Pennsylvania Department of Education, a multi-year evaluation of "students achieving standards" redeployment of I.T. hardware and software.
2. For Scholastic, Inc., national analysis of a technology-enriched literacy program for adolescents two or more years below grade level.
3. For the Milken Family Foundation, statewide multi-year analysis of the achievement outcomes of West Virginia's Basic Skills/Computer Education Program.
4. For Lightspan, Inc., multi-year multi-school analysis of the score gains and associated benefits of a school-home-school, "serious play" CD-ROM platform "curriculum of the home."
5. For Classwell Learning Group, Inc., an analysis of the outcomes of an online curriculum and related achievement.
6. For the US Department of Education, consulting assistance and trial demonstration of the measurement of student and other outcomes associated with learning technology.
7. For Leaming.com, a documentation study of the ISTE standards and achievement outcomes associated with an online curriculum.
8. For Homeroom.com, a yearlong analysis of score gains and other results from an online test preparation service.
9. For Sun Microsystems, a comparative analysis of the Total Cost of Ownership of thin client computing applications in California districts.
10. For the Gruss Life Monument Funds, an analysis of the outcomes of a computer delivered basic skills curriculum in selected Jewish day schools.
11. For the National Science Foundation, outcome documentation of "The Green Project," a curriculum developed to facilitate success in math and science among urban adolescent girls of color.
12. For the National Teachers Network, implementation and evaluation assistance in using telecommunications to connect teachers to each other for classroom improvement (now being extended by Citigroup to the United Kingdom and the Republic of Ireland).

Key Personnel

- Dale Mann, Ph.D., Managing Director, Interactive, Inc. and Professor, Organization and Leadership, Teachers College, Columbia University. Dr. Mann has been involved with school reform since the 1960's where his Washington service included work on the Elementary and Secondary Education Act and responsibility as Special Analyst for Education in the Executive Office of President Lyndon Johnson. Dr. Mann is author of books and articles on school reform including Making Change Happen? He is the founding chair of the International Congress for School Effectiveness, an organization with members from 65 countries focused on improving schools for the most needy children.

- Charol Shakeshaft. Ph.D., Managing Director, Interactive, Inc. and Professor of Foundations, Leadership, and Policy Studies at Hofstra University. Dr. Shakeshaft's specialties include research methods, using data for school decision-making, race and gender equity and school organization. She is an internationally recognized researcher in the area of gender and race patterns in educational delivery and classroom interactions. She serves as an expert witness and consultant in a number of legal proceedings on sexual harassment of students, and works with school districts to develop policies and practices that decrease peer harassment and bullying of all types among students. She is the author of *Women in Educational Administration* (5th printing) and *Sexual Violence in Schools* (Jossey-Bass, Fall 2001).

- Jonathan Becker. J.D., M.Ed., Research Director, Interactive, Inc. specializes in researching the equity effects of learning technology. He has been the principal investigator on analysis of student outcomes for: *Lightspan Achieve Now!*™; Scholastic, Inc.'s *READ 180*; AT&T Foundation-funded research on telecommunications and racial attitudes; and, the Arizona School Facilities Board. After graduating *cum laude* from Duke University with a B.A. in Public Policy Studies, he received a law degree and a masters degree in curriculum and instruction from Boston College Law School and the Boston College Graduate School of Education. Mr. Becker helped to conceive of and design the joint degree program in law and education at Boston College, and was the first student to complete the requirements for the program. Mr. Becker is completing a Ph.D. in Politics and Education at Teachers College, Columbia University.

- Kara Sweeney is a researcher at Interactive, Inc. and a graduate student at The State University of New York (Stony Brook campus), where she is enrolled in the M.A.T. in English program. She has earned a B.A. in English from Duquesne University and was an editor for an educational and children's book publisher.

APPENDIX B. REFERENCES

Berends, M. (2000) Teacher-Reported Effects of New American School Designs: Exploring Relationships to Teacher Background and School Context. Educational Evaluation and Policy Analysis. v22(1), 65-82.

Cooper, R. (1998) Socio-Cultural and Within-School Factors That Affect the Quality of Implementation of School-Wide Programs. Report No. 28.

Clark, D.L. (1984) Effective Schools and School Improvement: A Comparative Analysis of Two Lines of Inquiry. *Educational Administration Quarterly*. 20(3)41-68.

Darling-Hammond, L. (1990) Instructional Policy into Practice: The Power of the Bottom over the Top. *Educational Evaluation and Policy Analysis*, 12(3), 233-41.

Datnow, A. and Castellano, M. (2000) Teachers' Responses to Success for All: How Beliefs, Experiences, and Adaptations Shape Implementation. *American Educational Research Journal*, 37(3), 775-799.

Fullan, M.G. (1991) The New Meaning of Educational Change. New York: Teachers College Press

Gingiss, P.L. (1992) Enhancing Program Implementation and Maintenance through a Multiphase Approach to Peer-Based Staff Development. *Journal of School Health*, 62(5), 161-66.

Goldhaber, D.D. and Brewer, D.J. (2000) Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-45.

Henk, W.A. and Melnick, S.A. (1995) "The Reader Self-Perception Scale (RSPS): A new tool for measuring how children feel about themselves as readers." *The Reading Teacher*, 48(6).

Joyce, B., & Showers, B. (1988) Student achievement through staff development New York: Longman.

Mann, D., (ed.) (1978) Making Change Happen? New York: Teachers College Press.

Mann, D. (1978) The User-Driven System and Modest Proposal. T*eacher College Record*, 79(3).

McLaughlin, M.W. (1990) 'The Rand Change Agent Study Revisited: Macro Perspectives and Micro Realities", *Educational Researcher*, 19(9), 11-16.

Miles, K.H. and Darling-Hammond, L. (1998) Rethinking the Allocation of Teaching Resources: Some Lessons from High-Performing Schools. *Educational Evaluation and Policy Analysis*, 20(1), 9-29.

Rand Corporation. (1974) *Volume I: A Model of Educational Change*. Research Report No. R-1589/1-HEW. Santa Monica, CA.

Rand Corporation. (1975) *Volume II: Factors Affecting Change Agent Projects*. Research Report No. R-1 589/2-HEW. Santa Monica, CA.

Rand Corporation. (1975) *Volume III: The Process of Change*. Research Report No. R-1589/3-HEW. Santa Monica, CA.

Rand Corporation. (1975) *Volume IV. The Findings in Review*. Research Report No. R-1589/4-HEW. Santa Monica, CA.

Ryckrnann, et. al. (1987) 'The Survey of Achievement Responsibility."

Rosenholtz, 5. (1989). Teachers' Workplace: The Social Organization of Schools. New York: Longman.

For more information about READ 180,
please contact our office at 1-877-234-READ.

Item #: 418637